THE UNIVERSITY
of EDINBURGH

Exploring and exploiting the enzymes involved in tambjamine YP1 natural product biosynthesis

Piera Marchetti

A Thesis Submitted for the Degree of Doctor of Philosophy

The University of Edinburgh
2018

Lay Summary

Antibiotic resistance is a rapidly spreading global phenomenon causing common infectious diseases to become untreatable. It is paramount that new antibiotics are brought to the clinic to combat this problem, as without new discoveries all bacterial infections will become resistant. Many of the antibiotics currently in use, like penicillin, were isolated from bacteria, as they produce defensive molecules to gain a competitive advantage in their ecosystem. Therefore, a better understanding of the different molecules that bacteria produce and how they produce them is crucial for the discovery and development of new antibiotics as well as other medicines.

This thesis investigates tambjamine YP1, a molecule produced by a marine microorganism called *Pseudoalteromonas tunicata*. This molecule has both antibacterial and anticancer activity and could be developed for use in the clinic. Learning how this antibiotic is produced in the bacterium will allow us to harness this process and make new and improved tambjamines that have better and more specific activity. Therefore, this work explores the many different protein molecular machines (enzymes) that work together to make tambjamine YP1.

Tambjamine YP1 consists of two fragments that are joined together to form the final product. This thesis has elucidated the method of creating each one of these fragments and has explored the incorporation of new building blocks into the fragments to make novel molecules. Eventually, this research could be taken forward to produce new tambjamines to treat resistant bacterial infections.

Abstract

Natural products are secondary metabolites produced by many different organisms such as bacteria, fungi and plants. These biologically active molecules have been widely exploited for medicinal purposes generating a large proportion of the drugs in clinical use today. Tambjamine YP1 is an antimicrobial, antimalarial and cytotoxic bipyrrole natural product produced by the anti-biofouling marine microorganism *Pseudoalteromonas tunicata*. Its biosynthetic pathway is encoded in an operon (the *tam* cluster) consisting of 19 genes, 10 of which have proposed biosynthetic functions involved in bipyrrole formation and attachment of a fatty amine tail. However, details of the biosynthesis of these moieties have not been elucidated. This thesis presents studies on the characterization of four of the recombinant enzymes isolated from *Escherichia coli* (TamA, TamD, TamF and TamH) and identifies their roles in the biosynthetic pathway of tambjamine YP1.

Biosynthesis of the bipyrrole backbone is proposed to begin with the formation of the first pyrrole ring from L-proline, a common route to pyrrole biosynthesis in natural products. In contrast, the second pyrrole ring is assembled in an unusual linear manner by two didomain enzymes, TamF and TamD. Sequence analysis of TamF suggests it is a didomain fusion composed of a chain length factor (CLF) and a ketosynthase (KS). Similarly, TamD is also an unusual didomain fusion of an acyl carrier protein (ACP) and an α-oxoamine synthase (AOS). Together TamF and TamD produce the first bipyrrole intermediate in the tambjamine YP1 biosynthetic pathway, 4-hydroxy-2,2'-bipyrrole 5-methanol (HBM). Mass spectrometry (MS) analysis shows that TamF can be loaded with a cysteine-bound, β-ketopyrrole thioester and the TamD ACP domain can be converted to the malonyl-ACP form. TamF then catalyzes the Claisen-like condensation between the malonyl-TamD ACP and its β-ketopyrrole substrate to elongate the chain by two carbons. The resulting TamD ACP-bound diketopyrrole product is a substrate for the TamD AOS domain that catalyzes a second Claisen-like condensation with L-serine to release the product, HBM. This work defines a role for TamF and TamD in HBM biosynthesis and presents the first *in vitro* production and detection of this key bipyrrole intermediate by high performance liquid chromatography (HPLC) and MS analysis.

Attached to the tambjamine YP1 bipyrrole backbone is a fatty amine tail predicted to be derived from a C12 fatty acid. Previously, this moiety was proposed to be activated as a CoA thioester by a fatty acid CoA ligase (FACL) from outside the *tam* cluster. However, sequence

analysis of TamA, an enzyme in the *tam* cluster with no previously assigned function, suggested that it is an unusual didomain fusion of a fatty acid AMP ligase (FAAL) with a C-terminal ACP domain that could carry out this function. Surprisingly, MS analysis of recombinant TamA revealed the presence of bound C11 and C12 fatty acid adenylates, which can be transferred to the ACP domain upon attachment of a 4'-phosphopantetheine (4'-PP) to a conserved serine residue. An *in trans* acyl-transfer assay was developed using the recombinant TamA ACP domain and was used to screen fatty acid substrates with variable chain lengths (C2-C16). It was found that TamA can efficiently transfer several fatty acids (C6-C13) with the best substrate being the C12 acid. Contrary to the previous hypothesis, MS analysis showed that the C12 acylated TamA ACP domain acts as a substrate for TamH, a downstream enzyme in the pathway. This interesting didomain fusion contains thioester reductase (TR) and transaminase (TA) domains which work sequentially to produce the C12 amine from the TamA ACP-bound fatty acid thioester. This amine product can subsequently combine with an HBM-derived aldehyde to generate the mature tambjamine YP1 natural product. Furthermore, preliminary data on the use of TamA for the production of the recently discovered class of *N*-acyl histidine amides from *Legionella pneumophila* is presented.

In summary, key enzymes from the tambjamine YP1 pathway have been functionally characterized. As well as defining their roles in the pathway, this work also has potential impact on understanding the biosynthesis of other pyrroles (e.g. prodiginines) and polyketides. This work lays the foundations for engineering of the Tam proteins to produce novel tambjamine derivatives for biological testing.

Acknowledgements

I would like to thank Prof. Dominic Campopiano for not only giving me the opportunity to do this PhD but also for his unwavering enthusiasm for the project. He has been a great support and help over these four years and it has been a privilege to work in his lab. My thanks also go to Prof. Michael Burkart for hosting me in his lab at UCSD for three months and to his whole group who were so welcoming and helpful.

This work was supported by the Biotechnology and Biosciences Research Council (BBSRC) through the EastBio doctoral training programme which has not only allowed me to develop my scientific skills but also meet some incredible people. I am grateful to Dr. Stephen McMahon and Prof. Jim Naismith (formerly of the University of St. Andrews, now Oxford) for all their help in carrying out crystal trials. I would also like to thank Dr. Logan Mackay and SIRCAMS for their help trying to keep the MS working.

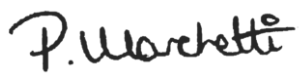
My thanks go to all the members of Lab 229 without whom my four years in Edinburgh would not have been the same. In particular, thank you to Annabel who has been with me every step of the way. It has been a privilege to have you as a friend as well as a colleague, talking through every problem and supporting me through the ups and downs. Thank you to Pete for not only holding the lab together and helping me whenever I needed it (including reading and correcting this thesis) but also for all the fun we've had, especially watching Scotland beat England at the rugby. Thank you to Guiomar who made joining the lab so easy, your warmth, kindness and love for nights out, made my first year so enjoyable. Thanks to Gary for always giving me a hug when I needed one, Mark for the best distractions, Alexis for the laughs, Jo for the kindness, Silvia for the incredible cakes and Ben for continuing to fly the Warwick flag. Thank you to Amanda, Stacie, Catherine, Chris, Bohdan, Lily, Alex, and Van and all our Masters and project students for all your help and for making it easy to come to work. A final thank you goes to Vero, Alisia, Sally and the Clarke group for making lunch time the best part of the working day.

A big thank you also has to go to my friends and family who have been there for me, supporting me and spurring me on throughout. It's hard to put into words exactly how much it means to me. To my mum and dad who have always believed in me and without whom I would not have made it here today. Thank you for supporting every step that I take, for helping me to achieve my dreams and pushing me to take risks and do the things that I love.

To Lindsey, thank you for being there every single day, to celebrate the good times and pick me up during the bad times (and make my dinner when I couldn't face it). Thank you for going above and beyond, for always saying yes to adventures and for making every day exciting, I can't wait for so many more.

Declaration

I, Piera Marchetti, declare that this thesis has been composed solely by myself and that it has not been submitted, in whole or in part, in any previous application for a degree or professional qualification. Except where otherwise stated by reference or acknowledgment, the work presented is entirely my own.



Piera Marchetti

The University of Edinburgh

11th March 2019

Abbreviations

4'-PP	4'-Phosphopantetheine
6xHis	Hexahistidine
AAT	Aspartate Aminotransferase
ACP	Acyl Carrier Protein
ACSM2A	Human Medium-Chain Acyl-Coenzyme A Synthetase
ALAS	Aminolevulinic Acid Synthase
AMP	Adenosine Monophosphate
ANL	Adenylation Enzyme
AONS	8-Amino-7-Oxononanoate Synthase
AOS	α-Oxoamine Synthase
APS	Ammonium Persulfate
AT	Acyltransferase
ATP	Adenosine Triphosphate
AUC	Area Under the Curve
BLAST	Basic Local Alignment Search Tool
bp	Base Pairs
C	Condensation Domain
CAPS	N-Cyclohexyl-3-Aminopropanesulfonic Acid
CAR	Carboxylic Acid Reductase
CHES	N-Cyclohexyl-2-Aminoethanesulfonic Acid
CLF	Chain Length Factor
CoA	Coenzyme A
CP	Carrier Protein
CV	Column Volume
DH	Dehydrogenase
DMAP	4-Dimethylaminopyridine
DNA	Deoxyribonucleic Acid
DMSO	Dimethyl Sulfoxide
dNTPs	Deoxynucleoside Triphosphate
EDC	1-Ethyl-3-(3-Dimethylaminopropyl) Carbodiimide
EDTA	Ethylenediaminetetraacetic acid

EIC	Extracted Ion Chromatogram
ESI-MS	Electrospray Ionisation - Mass Spectrometry
FAAL	Fatty Acid AMP Ligase
FACL	Fatty Acid Coenzyme A Ligase
FAD	Flavin Adenine Dinucleotide
FAS	Fatty Acid Synthase
FMN	Flavin Mononucleotide
GFC	Gel Filtration Chromatography
HBM	4-Hydroxy-2,2'-Bipyrrole-5-Methanol
HEPES	4-(2-Hydroxyethyl)-1-Piperazineethanesulfonic Acid
HMG-CoA	3-Hydroxy-3-Methyl-Glutaryl-Coenzyme A
HPLC	High Performance Liquid Chromatography
HTH	Helix-Turn-Helix
IMAC	Immobilized Metal Affinity Chromatography
IPTG	Isopropyl β-D-1-Thiogalactopyranoside
KBL	2-Amino-3-Ketobutyrate Coenzyme A Ligase
KS	Ketosynthase
LB	Liquid Chromatography
LC	Lysogeny Broth
MAS	Mycocerosic Acid Synthase
MAT	Malonyl-CoA:ACP Transacylase
MaxEnt	Maximum Entropy Algorithm
MBC	4-Methoxy-2, 2'-Bipyrrole Carbaldehyde
MBP	Maltose Binding Protein
mmCAR	*Mycobacterium marinum* Carboxylic Acid Reductase
MOPS	3-(*N*-Morpholino) Propanesulfonic Acid
NADH	Nicotinamide Adenine Dinucleotide
NADPH	Nicotinamide Adenine Dinucleotide Phosphate
NRP	Non-Ribosomal Peptide
NRPS	Non-Ribosomal Peptide Synthase
NSAID	Non-Steroidal Anti-Inflammatory Drugs
OAT	Ornithine Aminotransferase
ORF	Open Reading Frame

PBS	Phosphate Buffered Saline
PCP	Peptidyl Carrier Protein
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
PEA	Palmitoylethanolamide
PEG	Polyethylene Glycol
PEGMME	Polyethylene Glycol Monomethyl Ether
PK	Polyketide
PKS	Polyketide Synthase
PLP	Pyridoxal 5'-Phosphate
PMP	Pyridoxamine 5'-Phosphate
PP_i	Pyrophosphate
PPTase	Phosphopantetheinyltransferase
PTM	Post Translational Modification
RNA	Ribonucleic Acid
RONN	Regional Order Neural Network
SAM	*S*-Adenosylmethionine
SDR	Short Chain Dehydrogenases
SDS-PAGE	Sodium Dodecyl Sulfate - Polyacrylamide Gel Electrophoresis
SNAC	*N*-Acetylcysteamine Thioester
SOC	Super Optimal Broth
SPT	Serine Palmitoyltransferase
TA	Transaminase
TE	Thioesterase
TEMED	*N*,*N*,*N*',*N*'-Tetramethylethylenediamine
TEV	Tobacco Etch Virus
TFA	Trifluoroacetic Acid
TR	Thioester Reductase
Tris	Tris(Hydroxymethyl) Aminomethane
UV	Ultraviolet

Table of Contents

1 Introduction

1.1 Natural Products

Nature has produced a breadth of molecules that have had a significant impact on human history. Many plants, bacteria and fungi, as well as other organisms, produce small molecule secondary metabolites with potent biological activity. These molecules have been exploited for use in many industries but have most notably impacted on medicine. In the period between 1981 and 2014, 60% of newly approved small molecule drugs (n = 1211) were natural products, derivatives or mimics.[1] These figures are even higher for antibacterial and anticancer medications which rely heavily on natural product discovery. Traditional pharmaceutical drug discovery, utilizing high throughput screening of large libraries of synthetically accessible molecules, has a worryingly low hit rate of less than 0.001%. Comparing this figure to one example class of secondary metabolites, the polyketides (PKs), illustrates the utility of focusing on natural products; of the 7000 known structures of polyketides, more than 20 are used clinically, a hit rate of 0.3%.[2] However, though natural products lay the groundwork for potent biologically active compounds, many of these will require derivatization to improve activity, stability or uptake. Indeed, natural products have inspired many synthetic chemists to produce novel scaffolds or develop new chemistry to make these complex structures.[3,4] Therefore, there is an evident need for identification and understanding of new natural products for use as drugs or for inspiring new synthetic derivatives and mimics for the clinic.

Unfortunately, the rate of discovery of natural products has been decreasing over the past three decades with a concomitant decrease in drug approvals.[5] This reduction stems from the difficulty of identifying new biologically active molecules.[6] Conventional discovery relies on screening of extracts for killing activity against bacteria or fungi. However, as many species produce the same antimicrobial products, the risk of rediscovery and therefore duplication of effort is high. Supply of natural products for screening can also be a barrier, as the biological source of the molecule is required. Many of these sources are difficult to culture in the laboratory and may require sampling from their native environment. However, environmental or seasonal changes can significantly affect the production of these metabolites in the source organism.[6]

Novel approaches in the field are seeking to tackle the slowing rate of natural product discovery. Soil bacteria have traditionally been a fruitful source of biologically active molecules,[7] however only around 1% of these can be cultured in the laboratory. The advent of iCHIP,[8] a method of culturing bacteria using a diffusion chamber has radically improved this figure to 50%. Less well characterized environments like marine ecosystems are now being probed and exploited for novel chemistry.[9] Expanded access to genetic information with the improvement in speed and cost of sequencing technology has been a significant development for the field.[10,11] We now have a much greater understanding of the enzymes involved in natural product biosynthesis and we can use this knowledge to probe genomes for the production of unique secondary metabolites. Expanding our understanding of natural product biosynthetic genes and enzymes is of utmost importance to facilitate their discovery.

1.2 Pyrrole Natural Products

Nitrogen heterocycles are the building blocks of nature, appearing in many vital molecules including deoxyribonucleic acid (DNA), ribonucleic acid (RNA) and proteins.[12] Pyrrole moieties are particularly ubiquitous and constitute part of many interesting natural products. The versatile pyrrole motif is planar and electron rich, resulting in its ability to form hydrogen bonds, chelate metal ions and participate in π-stacking interactions.[13] One of the most notable pyrrole containing compounds is heme, the tetrapyrrole porphyrin cofactor (Figure 1.1). This structure consists of four pyrrole rings attached in a larger, planar system. The ring system is able to chelate iron to carry out various activities including electron shuttling in cellular respiration, oxygen binding and oxygenation.[14] Pyrrole is also a key component in the indole structure of tryptophan, where it forms an integral part of the majority of proteins.

Pyrrole natural products constitute a large number of biologically active molecules that have potential for use as drugs (Figure 1.1). Clorobiocin is a bacterial DNA gyrase inhibitor effective against *Escherichia coli* whose pyrrole ring is essential for activity;[15] pyoluteorin induces Mcl-1 degradation and apoptosis in cancer cells;[16] prodigiosin is both an antibacterial and anticancer agent able to intercalate and cleave DNA;[17] the makaluvamines are DNA topoisomerase II inhibitors with anticancer activity[18] and the newly isolated heronapyrroles have strong Gram-positive antibacterial activity.[19] Though these natural products have yet to make it to the clinic they have inspired several novel derivatives with enhanced activity.[16,20,21]

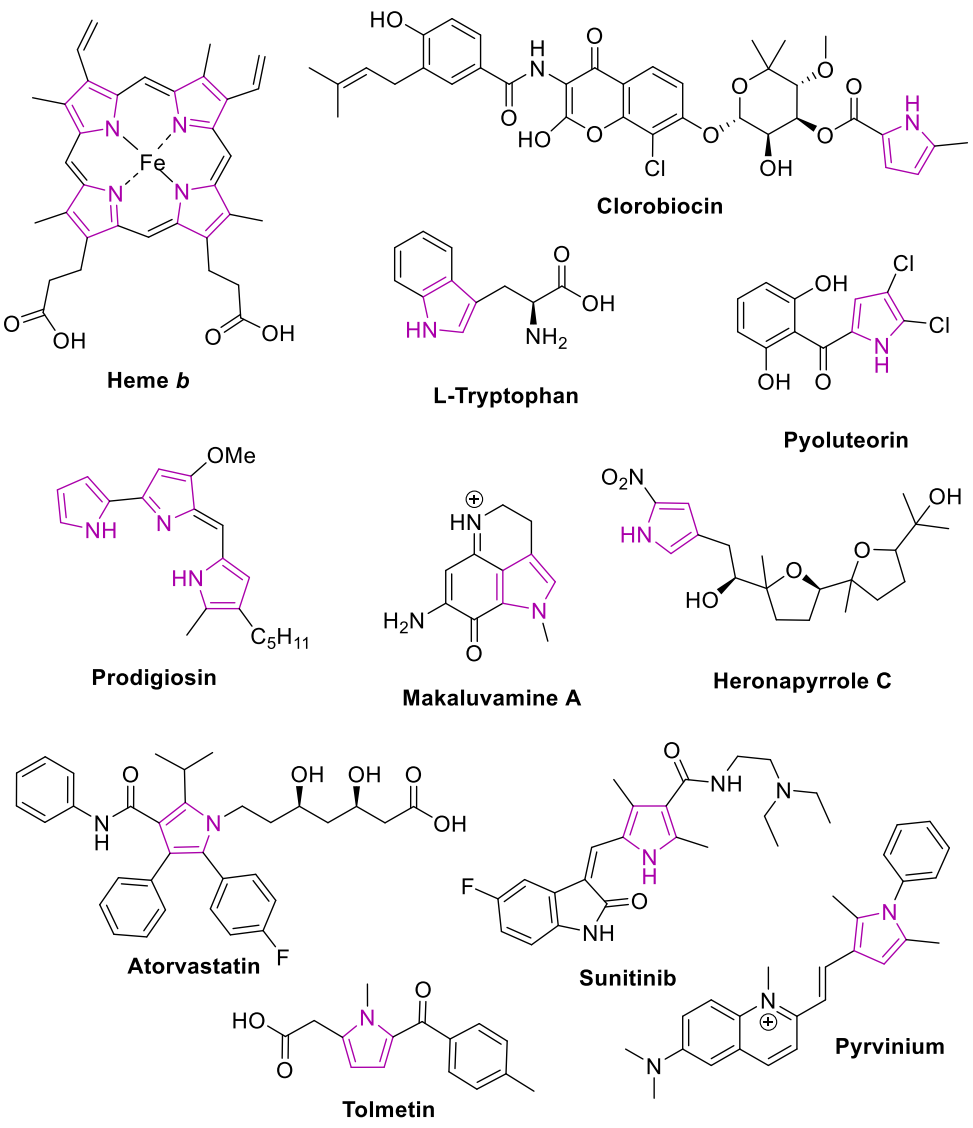


Figure 1.1: Molecules with biological activity containing pyrrole groups (purple). Heme *b* and L-tryptophan form essential components of hemoglobin and proteins. Clorobiocin, pyoluteorin, prodigiosin, makaluvamine A and heronapyrrole C are all pyrrole containing natural products with potential for use as drugs. Atorvastatin, tolmetin, sunitinib and pyrvinium are FDA or EU approved drugs containing pyrrole moieties.

The potential of pyrrole containing molecules is exemplified by the incorporation of pyrrole into several approved drugs with a wide variety of indications, including atorvastatin, tolmetin, sunitinib and pyrvinium (Figure 1.1). Atorvastatin is an inhibitor of the enzyme 3-hydroxy-3-methyl-glutaryl-coenzyme A (HMG-CoA) reductase which catalyzes the rate-limiting step of cholesterol biosynthesis in the liver. This causes the liver to uptake cholesterol from the blood, lowering its concentration and decreasing the patient's risk of

cardiovascular disease.[22] Though tolmetin's mode of action is unknown, it is a non-steroidal anti-inflammatory drug (NSAID) that has been successfully used for the treatment of pain and inflammation in arthritis and other diseases.[23] Sunitinib is another enzyme inhibitor which binds to receptor tyrosine kinases involved in tumor cell angiogenesis and proliferation. It has been approved for use in renal cell carcinoma and gastrointestinal stromal tumour.[24] The final pyrrole-containing drug is pyrvinium, which is used as an antihelminthic[25] and has also been employed to treat infection of the parasite *Cryptosporidium.*[26] These molecules illustrate that pyrrole rings are distinctly important elements in biologically active molecules and natural products. Identification of new pyrrole containing natural products and elucidation of their biosynthesis will be an important step in developing new medications.

1.3 Tambjamines

1.3.1 Structure and Activity

One interesting family of pyrrole natural products containing two rings in a core 2,2'-bipyrrole structure are the tambjamines, a family of yellow pigments (Figure 1.2). Their bipyrrole core is similar to that of a well characterized natural product family – the red-coloured prodiginines,[27] which share the 4-methoxy-2,2'-bipyrrole carbaldehyde (MBC, Figure 1.2) backbone. Tambjamines consist of the MBC aldehyde (sometimes containing bromine substituents) attached to a fatty amine tail *via* an enamine linkage.

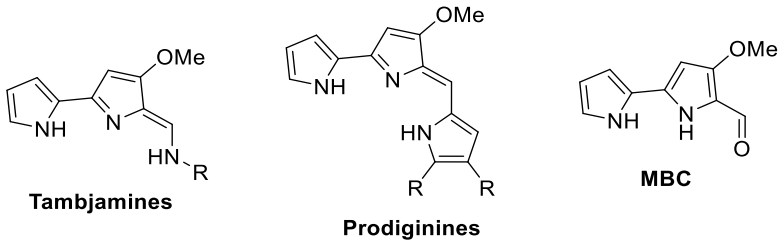


Figure 1.2: General structures of the tambjamines (R = hydrocarbons between C2 and C12), the prodiginines (R = various hydrocarbons and cyclisations) and 4-methoxy-2,2'-bipyrrole carbaldehyde (MBC), the bipyrrole core of these molecules.

In contrast, the prodiginines consist of a third pyrrole ring attached to the bipyrrole backbone *via* a methylene bridge, resulting in a conjugated tripyrrole system. The third ring in the

prodiginines is generally substituted with hydrocarbon chains that can also be cyclized, as in the case of streptorubin B.[27]

The first tambjamine molecules were isolated in 1982 by Carté and Faulkner who elucidated the structures of tambjamines A-D[28] (Figure 1.3). These molecules originated from the marine environment in the *Tambja* genus of nudibranchs and their bryozoan prey *Sessibugula translucens*. The tambjamines appeared to be utilized as defensive chemicals to ward off predators like the carnivorous nudibranch *Roboastra tigris*.[29] A few years later, tambjamines E and F were identified in another marine feeding association between *Nembrotha* nudibranchs and the *Atapozoa* ascidians.[30] These molecules play the same defensive role in these food chains, also deterring predatory fish.[31] Four other tambjamines (G-J) were subsequently extracted from a different bryozoan, *Bugula dentata*.[32] These new molecules showed strong biological activity in a killing assay of brine shrimps, suggesting the same defensive function.

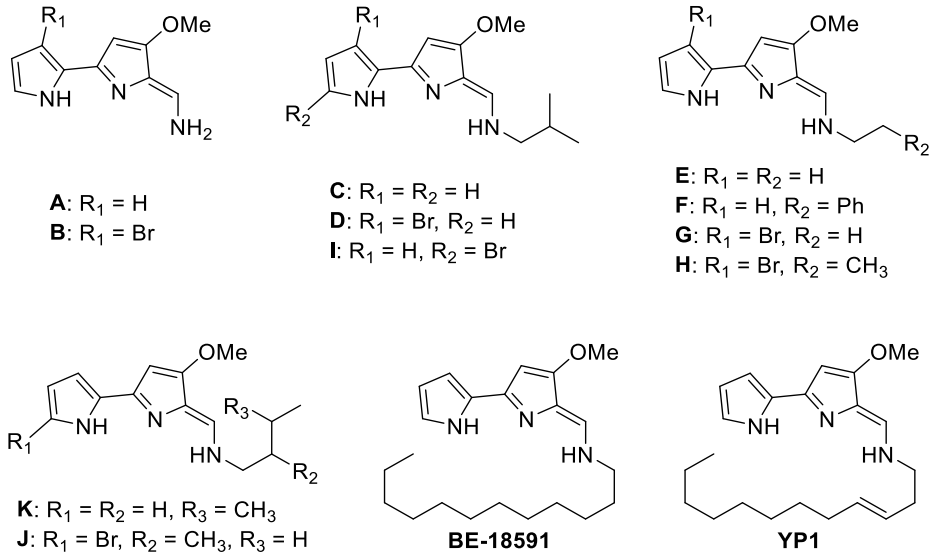


Figure 1.3: Structures of the naturally occurring tambjamines A-K, YP1 and BE-18591 isolated from nudibranchs, bryozoans, ascidians and bacteria.

Only one tambjamine, named BE-18591, has thus far been identified outside the marine environment, isolated from a Japanese plant dwelling *Streptomyces sp.*[33,34] BE-18591 has a much longer fatty acid chain than the tambjamines previously characterized, consisting of 12 carbons (Figure 1.3). The most recent naturally occurring tambjamine identified is tambjamine YP1 which was isolated from the marine bacterium *Pseudoalteromonas tunicata*.[35] This molecule has a very similar structure to BE-18591 but contains a double bond

at position 3 of the hydrocarbon tail. YP1 also has deterrent properties as it is responsible for preventing the biofouling of its host organism, the common alga *Ulva australis*.

Since the tambjamines are thought to be involved in chemical defense and also show killing activity, these molecules have been explored for their potential therapeutic value. They display potent antimicrobial activity against the pathogens *Candida albicans*, *E. coli, Vibrio anguillarum* and *Malassezia furfur*[28,36] as well as cytotoxic activity against mammalian cancer cells.[36–41] This activity arises from two mechanisms of action. The tambjamines bind DNA in the minor groove and are able to intercalate between adenine-thymine base pairs (bp).[36–38] They subsequently catalyze copper dependent oxidative cleavage of the DNA leading to DNA fragmentation and apoptosis. Though the naturally occurring tambjamines have not shown particular selectivity for cancer cells, they have inspired new tambjamine analogues with improved activity and selectivity.[40] The second mechanism of action of the tambjamines is transmembrane anion exchange.[42] These molecules can transport bicarbonate and chloride ions across cell membranes which is likely to play a role in their antimicrobial and cytotoxic activity. However, this activity could also be tuned for therapeutic use in the treatment of channelopathies. The Quesada group has shown that increasing lipophilicity of the tambjamine hydrocarbon tail improves anion transport up to a chain length of 9-carbons.[43] Unnatural analogues have also been designed with different aromatic substituents that have even greater anion transport activity.[42,44] Finally, novel tambjamines have also been created with excellent activity against malaria.[45] The best performing molecule, KAR425, a tambjamine with a saturated 7-membered ring in place of its hydrophobic tail, showed greater efficacy than any other tambjamine or prodiginine analogue and was curative in mice at 80 mg/kg in a single dose. Taken together, these results suggest that tambjamine molecules have very promising potential as medicinal candidates.

1.3.2 Tambjamine YP1

Tambjamine YP1 is the most recently isolated natural tambjamine, extracted from the marine bacterium *P. tunicata* by the Kjelleberg group.[46] *P. tunicata* was discovered both on the surface of tunicates and alga where it competed against other biofouling organisms for space and nutrients.[47,48] Its success in these environments relies on the production of a number of

secondary metabolites with inhibitory activity, including the well-known antibiotic violacein,[49] along with tambjamine YP1.

In order to better understand tambjamine YP1 biosynthesis, the Kjelleberg group identified and cloned the gene cluster responsible for its production in *P. tunicata.*[50] To date, this is the only known tambjamine producing gene cluster. The group generated a *P. tunicata* genomic library of ~35 kb fragments expressed in *E. coli* and determined which clones produced tambjamine by fungal growth inhibition assay. Three clones were able to biosynthesize the yellow pigment and inhibit fungal growth. Analysis of these clones revealed an overlapping region of 21 open reading frames (ORFs), 19 of these forming a 26 kb unidirectional cluster (Figure 1.4). Mutants of the genes in the cluster resulted in abolition of pigment production and fungal growth inhibition. Therefore, this cluster was designated as the *tam* cluster.

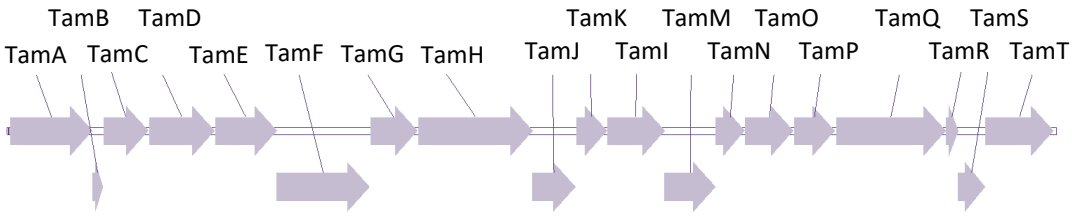


Figure 1.4: The *tam* gene cluster responsible for the biosynthesis of Tambjamine YP1 containing 19 open reading frames (ORFs), annotation according to Burke *et al.*[50]

Each of the 19 ORFs from the cluster were subjected to Basic Local Alignment Search Tool (BLAST)[51] analysis to identify the conserved protein domains. The genes were also compared to the *red* and *pig* clusters of *Serratia marcescens* and *Streptomyces coelicolor.*[27,52] These clusters produce the tripyrrolic prodiginines with the same core bipyrrole structure. Table 1.1 shows the predicted functions of the proteins as well as their size, GenBank access codes and their homologues from the prodiginine biosynthetic gene clusters.

The conserved domain analysis is immediately interesting as 4 of the 19 ORFs in the cluster encode fused didomain enzymes that have multiple functions (TamA, TamD, TamF and TamH). Utilizing the predicted domains of each protein, as well as their homology to the closely related clusters, Burke *et al.*[50] postulated the biosynthesis of tambjamine YP1. This biosynthetic pathway is presented in Figure 1.5 with some updates, including the correction of chemical errors. The pathway appears to be a convergent biosynthesis consisting of two

fragments, the bipyrrole MBC and a long chain fatty amine, which are condensed to form the final product.

Gene	GenBank Code	Homologues	Predicted Size (kDa)	BLAST Conserved Domains
tamA	EAR29369	-	75	Fatty acid AMP-ligase/ACP
tamB	EAR29368	pigG/redO	10	PCP
tamC	EAR29367	redG	43	Rieske oxidase
tamD	EAR29366	pigH/redN	58	8-Amino-7-oxononanoate synthase
tamE	EAR29365	pigI/redM	56	Amino acid adenylating enzyme
tamF	EAR29364	pigJ/redX	81	β-ketoacyl synthases (2 domains)
tamG	EAR29363	pigA/redW	43	Acyl CoA dehydrogenase
tamH	EAR29362	pigE	104	TA/Acyl-ACP reductase
tamI	EAR29359	-	50	ABC transporter permease
tamJ	EAR29361	pigM/redV	40	FMN nitroreductase
tamK	EAR29360	-	29	Periplasmic sorting protein
tamM	EAR29358	-	45	ABC transporter permease
tamN	EAR29357	-	26	ABC transporter ATPase
tamO	EAR29356	-	45	Unknown
tamP	EAR29355	pigF/redI	37	SAM-dependent methyltransferase
tamQ	EAR29354	pigC/redH	100	Phosphoenolpyruvate synthase
tamR	EAR29353	pigK/redY	12	Unknown
tamS	EAR29352	pigL/redU	25	PPTase
tamT	EAR29351	-	62	Acyl-CoA dehydrogenase

Table 1.1: The *tam* cluster genes with their GenBank codes, homologues from the *pig* and *red* clusters, predicted protein size and conserved domains; updated from Burke *et al.*[50]

MBC is a common intermediate of the prodiginine and tambjamine natural products and the biosynthesis of this molecule has been explored and reviewed.[27,52] Therefore, the predicted pathway to MBC production in *P. tunicata* is largely based on the information obtained from the *pig* and *red* clusters. Generation of the bipyrrole fragment is predicted to begin with TamE, a putative adenylation enzyme (ANL) which is specific for L-proline. This enzyme activates L-proline using adenosine triphosphate (ATP) to form a proline adenylate intermediate which can then be transferred onto TamB, a peptidyl carrier protein (PCP). PCPs and acyl carrier proteins (ACPs) are non-catalytic and serve to anchor natural product intermediates while they are being extended or modified.[53] The substrates are bound to their

4'-phosphopantetheine (4'-PP) post-translational modification as thioester intermediates. Once the proline residue is bound, the PCP conveys it to the active site of TamG, a putative flavin adenine dinucleotide (FAD)-dependent dehydrogenase. TamG is likely to carry out a two-step oxidation of the proline heterocycle to the aromatic pyrrole. The β-ketopyrrole is subsequently handed off to TamF, a ketosynthase (KS) / chain length factor (CLF) didomain protein. TamD, another interesting didomain enzyme works with TamF to effect chain extension. TamD contains an ACP which becomes loaded with a malonyl group by an as yet unknown mechanism and an α-oxoamine synthase (AOS) domain. TamF performs a Claisen-like condensation of the malonyl-TamD ACP with its β-ketopyrrole intermediate. This extends the chain by two carbons and the TamD ACP thioester can now act as a substrate for the TamD AOS domain. This domain is expected to bind L-serine and carry out a pyridoxal 5'-phosphate (PLP)-dependent decarboxylative condensation to release an α-oxoamine. This linear molecule can spontaneously cyclise to produce 4-hydroxy-2,2'-bipyrrole-5-methanol (HBM), the first bipyrrole intermediate in the pathway. Two additional tailoring reactions are required to reach the final bipyrrole fragment: oxidation of the alcohol group originating from the serine side chain by TamJ and methylation of the 4-hydroxy group by TamP, an *S*-adenosyl methionine (SAM)-dependent methyltransferase.

The evidence for this predicted MBC pathway is based on knockout experiments in the *red* cluster of *S. coelicolor* and *in vitro* studies of the Pig enzymes from *S. marcescens*. Knockouts of the *red* genes showed that RedM, W, X and N (homologues of TamE, G, F and D, the adenylation enzyme, FAD dehydrogenase, KS and AOS respectively) are necessary for the formation of the prodiginines produced by *S. coelicolor.*[54] However, feeding of these mutants with synthetic MBC restored production of the natural products, showing that these enzymes are responsible for biosynthesizing this fragment. The Challis group also demonstrated that feeding of a small molecule pyrrole-PCP mimic to RedM and RedW mutants (the adenylation and FAD dehydrogenase domains respectively) restored natural product production.[54] However, pyrrole carboxylic acid (an oxidized proline molecule) was unable to restore production, illustrating that this intermediate is never released. This strongly suggests that the pyrrole ring is transferred directly from the PCP to the KS. In the same year, the Walsh group expressed and characterized the proteins involved in HBM biosynthesis in the *pig* cluster.[55] They were able to detect all enzyme bound intermediates from L-proline to the diketopyrrrole-ACP. However, they were unable to identify the final HBM product, thus, it has never been confirmed as the first bipyrrole intermediate in the pathway.

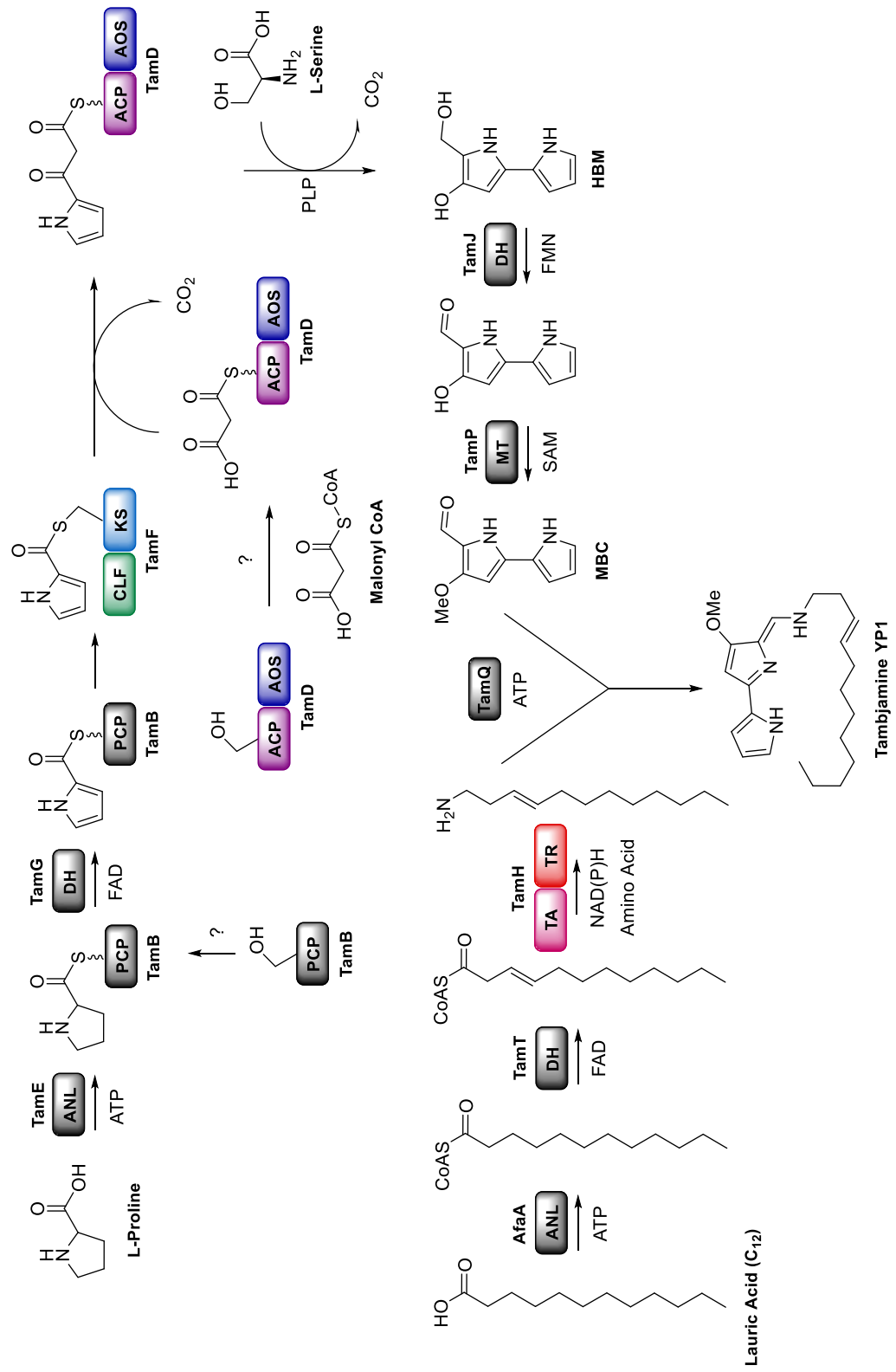


Figure 1.5: Predicted biosynthetic pathway to tambjamine YP1 from L-proline, malonyl-CoA, L-serine and lauric acid in *P. tunicata,* adapted and corrected from Burke *et al.*[50]

Burke *et al*.[50] postulated that the tambjamine YP1 amine tail originates from lauric acid, C12 carboxylic acid. This pathway would be initiated by AfaA, an enzyme from outside the *tam* cluster that has homology to fatty acid coenzyme A ligases (FACLs). When this gene is knocked out in *P. tunicata*, yellow pigment formation is abrogated, so it was predicted that AfaA produces a coenzyme A (CoA) bound lauric acid *via* an adenylate intermediate. The CoA derivative can then be dehydrogenated by the FAD-dependent dehydrogenase TamT, to introduce the 3,4 double bond. Finally, the bifunctional TamH can produce the amine by reductive amination. This unusual enzyme consists of an N-terminal transaminase (TA) and a C-terminal thioester reductase (TR). The reductase domain should reduce the CoA thioester to produce free CoA and C12 aldehyde using nicotinamide adenine dinucleotide (or phosphate, NAD(P)H). This aldehyde can subsequently be transaminated by the TamH N-terminal domain to produce the final amine fragment. However, none of these biosynthetic transformations have been probed as this pathway is not shared with prodiginine biosynthesis.

After completion of the building blocks, the final step is condensation of the two fragments by TamQ to form tambjamine YP1. The condensation enzymes of the prodigiosin pathways have been well studied and have been exploited for the formation of novel natural products as they accept a variety of substrates.[21,56,57] As TamQ is similar to these enzymes, it is possible that it will also accept a range of building blocks to produce new and improved tambjamines.

1.4 Non-Ribosomal Peptide (NRP) and Polyketide (PK) Biosynthesis

1.4.1 Biosynthetic Pathway

Many of the predicted genes in the tambjamine YP1 pathway share similarity to genes involved in the biosynthesis of two groups of natural products, the PKs and the non-ribosomal peptides (NRPs). A marked increase in the understanding of the genomics of these natural products arrived with the cloning of 6-deoxyerythronolide B synthase (DEBS), an enzyme involved in erythromycin biosynthesis by both Leadlay[58] and Katz[59] in the early 1990s. This introduced us to the 'assembly line' biosynthesis of PKs by PK synthases (PKSs) and shortly after a similar arrangement was described for NRP synthases (NRPSs).[60] In these examples, large polypeptides containing several functional domains grouped into modules

act as a conveyor belt for their natural products.[61,62] A collinearity was found between the number of modules and their specific domains and the number of building blocks incorporated and how they are modified.[10]

Both systems have a very similar assembly line which is analogous to the biogenesis of fatty acids in primary metabolism by the fatty acid synthase (FAS).[63,64] They begin with activation of a substrate and transfer to a carrier protein (CP). Elongation occurs by activation of a second substrate and Claisen-like condensation between the second and the first. This cycle can be repeated several times before chain termination, releasing the product (Figure 1.6).

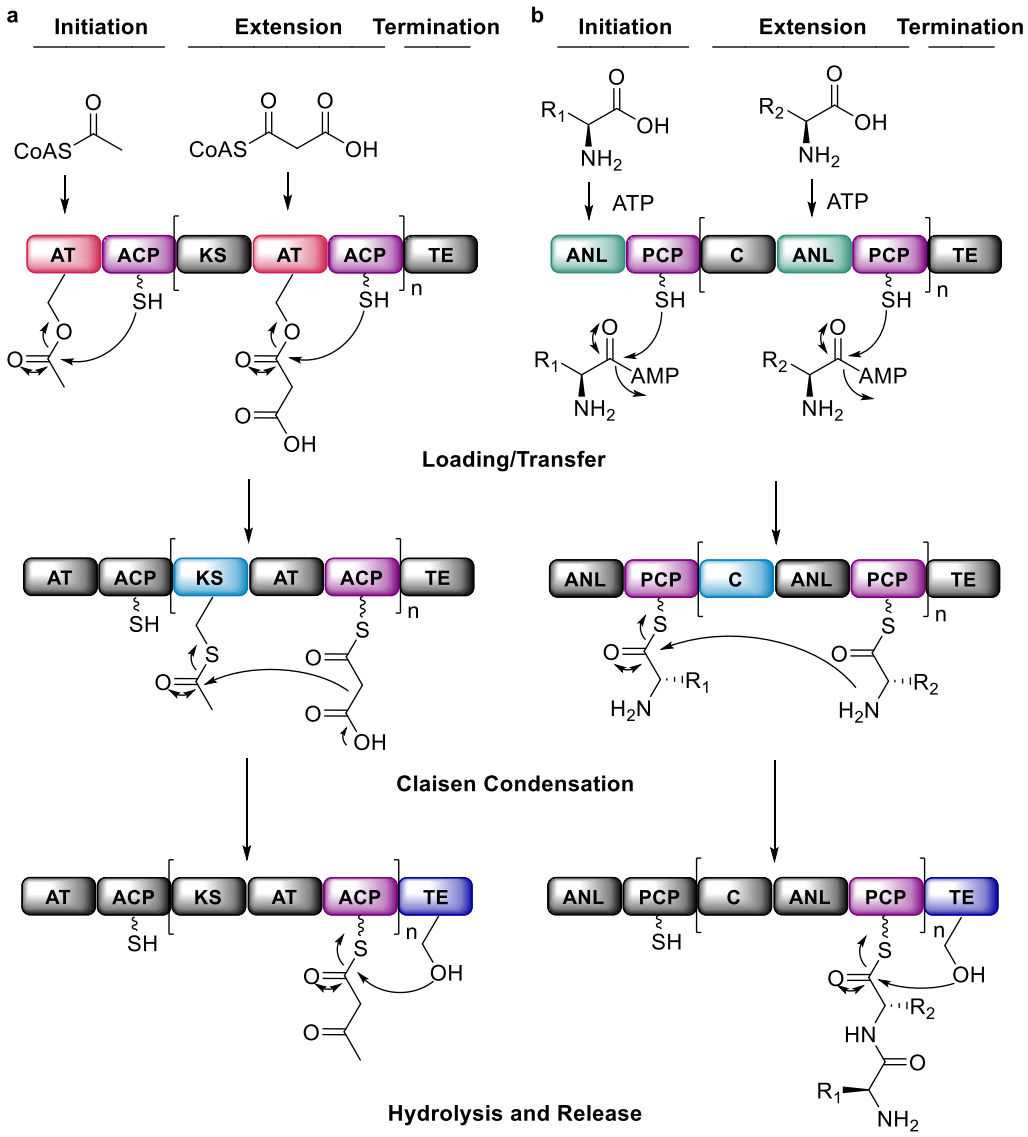


Figure 1.6: General biosynthetic route of (**a**) polyketide (PK) and (**b**) non-ribosomal peptide (NRP) natural products adapted from Weissman.[61]

The PKS carries out highly similar reactions to the FAS utilizing the same set of starter and elongation units, usually acetyl-CoA and malonyl-CoA respectively. The PKS begins with an initiation module containing an acyltransferase (AT) domain and an ACP which load up the starter unit for the elongation module. The initiation domain AT transfers the starter unit from an activated CoA substrate onto its conserved serine residue before delivering it to the 4'-PP modified ACP. Transthioesterification can now occur from the ACP to the cysteine residue of the elongation module KS. This module also contains an AT domain that loads the second building block, containing a free carboxylic acid, onto the second ACP. Claisen-like condensation is catalyzed by the KS which decarboxylates the second building block to form a carbanion to attack the KS bound starter unit, increasing the chain length by two carbons. The first ketone can now be modified, if required, by reduction to an alcohol by a ketoreductase, dehydration to a double bond by a dehydratase and further reduction to the single bond by an enoyl reductase. Any or all of these modifications can be performed, increasing the variability of PKS products. This is in contrast to fatty acid biosynthesis where all of these steps would be performed at every stage to yield a saturated hydrocarbon. The variability of PKs can also be increased with the incorporation of alternative starter or extender units such as propionyl-CoA or methylmalonyl-CoA. After extension by a defined number of building blocks, the chain is terminated and released, most commonly by a thioesterase (TE). This enzyme attacks the ACP thioester forming an ester on its conserved serine residue. This ester is subsequently hydrolyzed to release the polyketide chain with a terminal carboxylic acid.

Rounds of extension are carried out in a similar way in the NRPS, where an amino acid is activated by an ANL domain using ATP to produce an amino acid adenylate. This adenylate can be attacked by the thiol of the PCP to form the activated thioester in the initiation module. A second ANL domain activates a second amino acid in the elongation module and the condensation (C) domain catalyzes the Claisen-like condensation between the two. Though condensation domains carry out a reaction analogous to the KS, they do this in a slightly different manner. The starter unit is not transferred to the condensation domain, instead, this enzyme catalyzes the extension between the two amino acids bound to each PCP. Similarly, this can occur several times before chain termination by a TE to release the peptide.

These two modular constructs utilize very similar methodology, including CPs and condensation reactions. For this reason they can also form hybrid synthases allowing both PKS and NRPS modules to work together to create natural products containing elements of each type.[65] Further research into the area has also identified synthases with only one module that incorporate more than one building block, known as iterative synthases.[66] Type II PKSs have also been described, which consist of discrete enzymes containing only one catalytic domain per polypeptide, similar to the type II FASs of bacteria.[67] These proteins are usually also iterative, with the ACP moving between its catalytic partners several times. The predicted tambjamine YP1 pathway is comprised of elements from both PKS and NRPS biosynthesis as it incorporates proline and malonyl units in a type II style system with only one or two functional domains per polypeptide.

1.4.2 Carrier Proteins (CPs)

The overview of FAS, PKS and NRPS biosynthesis, as well as the predicted pathway to the biosynthesis of tambjamine YP1 has shown that CPs are the backbone of natural product pathways, shuttling the growing chain between the many different catalytic partners they encounter.[68] These systems have similar domain organization, but FAS and PKS are evolutionarily distinct from NRPS suggesting that CPs are employed in many different carboxylate-utilizing reactions.[69] These proteins are small, at less than 100 amino acids, they have a molecular weight between 8 and 10 kDa.[53]

CPs must be post-translationally modified with a 4'-PP arm derived from CoA to convert them from *apo*- to *holo*-form. This modification is necessary for CPs to attach their substrates as activated thioesters and is carried out by the phosphopantetheinyltransferases (PPTases, Figure 1.7).[69] The 20 Å length of the 4'-PP arm also allows it to stretch directly into the active sites of partner enzymes.[70]

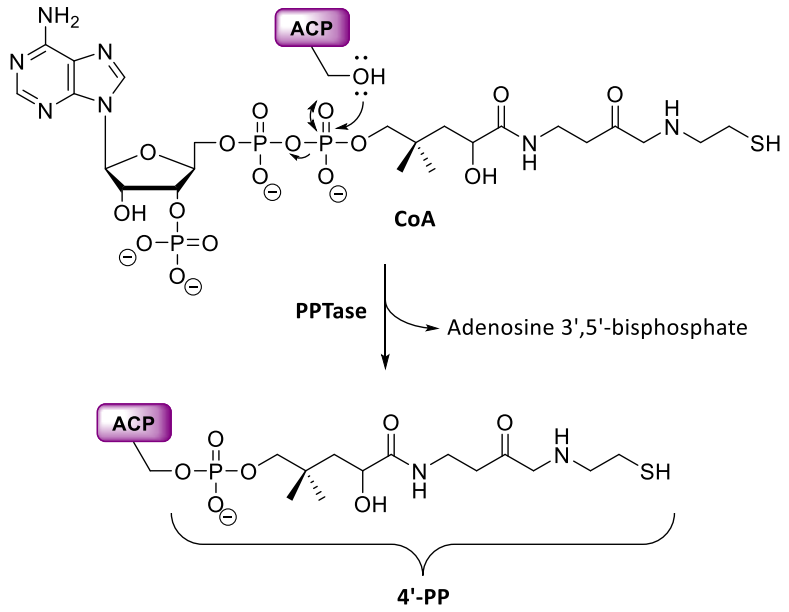


Figure 1.7: Activation of carrier proteins (CPs) by transfer of 4'-phosphopantetheine (4'-PP) from coenzyme A (CoA) onto a conserved serine residue by a phosphopantetheinyltransferase (PPTase).[71]

The ACP was first identified in the bacterial type II FAS where it transports developing fatty acids between each discrete catalytic enzyme. The first defined structure of an ACP was the solution NMR structure of the *E. coli* FAS ACP[72] which has since been crystallized in various forms.[73,74] The *E. coli* ACP structure illustrates the now canonical ACP fold (Figure 1.8a) consisting of three parallel α-helices (helices I, II and III) with a fourth shorter anti-parallel helix (helix IV) running across these. The three parallel helices contain mainly hydrophilic residues on the solvent exposed surface and hydrophobic residues facing inwards, forming a hydrophobic cleft in the center of the protein. The helices are connected by flexible loop regions with loop I and helix II showing the highest flexibility.[75] The dynamic structure of these proteins has resulted in the difficulty of obtaining crystal structures of ACPs, so NMR has often been used in its place.[76]

The 4'-PP modification of CPs takes place at the top of Helix II, on a serine residue in the conserved DSX motif where X is usually leucine or another small hydrophobic amino acid.[77] This modification confers greater stability to the ACP structure[77,78] causing it to contract slightly.[79] The *holo*-forms of ACPs are mainly stabilized by hydrophobic interactions of the 4'-PP arm with the protein structure[53] including the X amino acid from the DSX motif.[79]

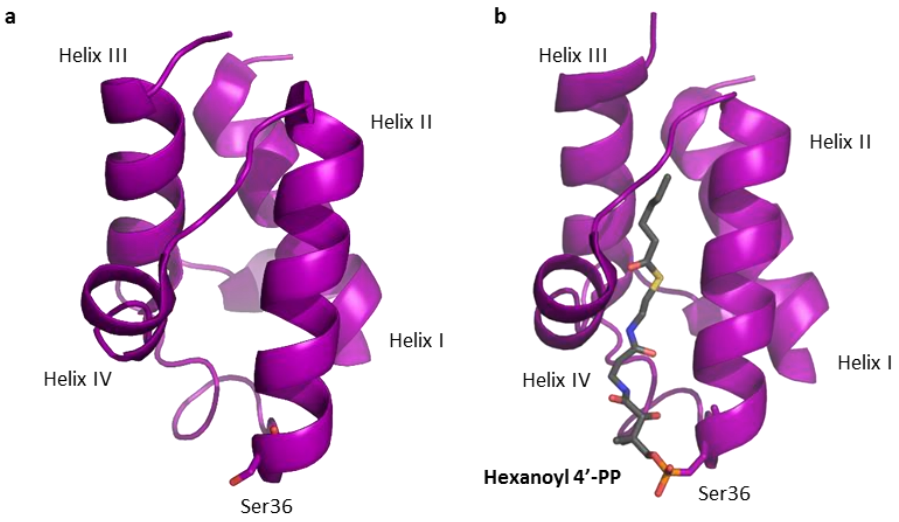


Figure 1.8: Crystal structures of *E. coli* acyl carrier protein (ACP) showing the conserved four-helix structure (**a**) in its *apo*-form, protein data bank (PDB) code: 1T8K[73] and (**b**) in its 4'-phosphopantetheine (4'-PP) modified *holo*-form with an attached hexanoyl substrate (grey), PBD code: 2FAC.[74] The conserved serine residue (Ser36) that becomes post-translationally modified is also shown as sticks in each structure.

ACPs are additionally stabilized by acylation during biosynthesis when the hydrophobic cleft contracts even further.[80] Type II ACPs have been shown to sequester their 4'-PP arms containing acyl substrates to protect the thioester from hydrolysis (Figure 1.8b).[81,82] The chains are bound in the central hydrophobic cavity of the protein formed predominantly by Helices II and IV[53] and can expand to accommodate longer chains.[74] When the ACP is required to deliver its substrate to a catalytic partner the 4'-PP arm 'flips out' of the hydrophobic cavity into the active site of the partner enzyme.[70,83] This ACP-partner interaction is very specific, exchanging ACPs between catalytic systems rarely results in full activity and often shows no activity at all. Helix II, containing the 4'-PP modification, is mainly responsible for mediating this interaction and thus is termed the 'recognition helix'.[84] This part of the protein contains conserved acidic residues which form specific contacts with the partner enzymes.[85] Although specific, these interactions are also transient as the ACP must be able to move on and off its respective partner enzymes with ease.[84]

The majority of ACP structural data has been derived from type II systems. However, more recently structures of type I ACPs that form part of megasynthases have also been determined. These proteins are very similar in structure but there are some differences

between the two, as protein-protein interactions are less important for the type I ACPs whose partners are generally fused.[69] This results in the type I proteins being less acidic than their type II counterparts.[86] Additionally, type I ACPs do not sequester their substrates in hydrophobic pockets as they are less likely to suffer from hydrolysis.[87]

1.4.3 Phosphopantetheinyltransferases (PPTases)

CPs require post-translational modification with a 4'-PP arm in order to transport biosynthetic intermediates. The enzymes responsible for this modification are the PPTases which transfer 4'-PP from CoA onto the conserved serine residue of a CP (Figure 1.7). These proteins fit into three families:[88] the AcpS family, the Sfp family and the megasynthase family. The AcpS and Sfp families have been most widely researched and are much more common than PPTases as part of megasynthases.[89] As it is a stand-alone protein, the predicted PPTase in the tambjamine YP1 pathway must be similar to AcpS or Sfp. Though it is currently unclear which family this protein belongs to, it is likely to be an Sfp type protein as these are more commonly found in secondary metabolite biosynthesis.[71]

AcpS is the *holo*-ACP synthase for the type II FAS ACP from *E. coli* and other bacteria. The first purification of *E. coli* AcpS showed that its activity relies upon divalent metal ions, either Mg^{2+} or Mn^{2+}.[90] It has some substrate promiscuity for other similar type II ACPs but is not able to modify any type I ACPs including *E. coli* EntF from the enterobactin synthase.[91] The first structure of the *B. subtilis* enzyme indicated that it is a homotrimer with three active sites at each interface between the monomers (Figure 1.9a).[85] The Mg^{2+} ion is responsible for coordinating the diphosphate of CoA in the active site and stabilizing the anion during attack of the ACP serine.

In contrast, Sfp is a PPTase from secondary metabolism responsible for loading the surfactin synthase PCP in *B. subtilis.* Its discovery led to the realization that different biosynthetic pathways require different PPTases.[71] This enzyme was also crystalized, revealing an alternative structure to the AcpS family (Figure 1.9b).[92]

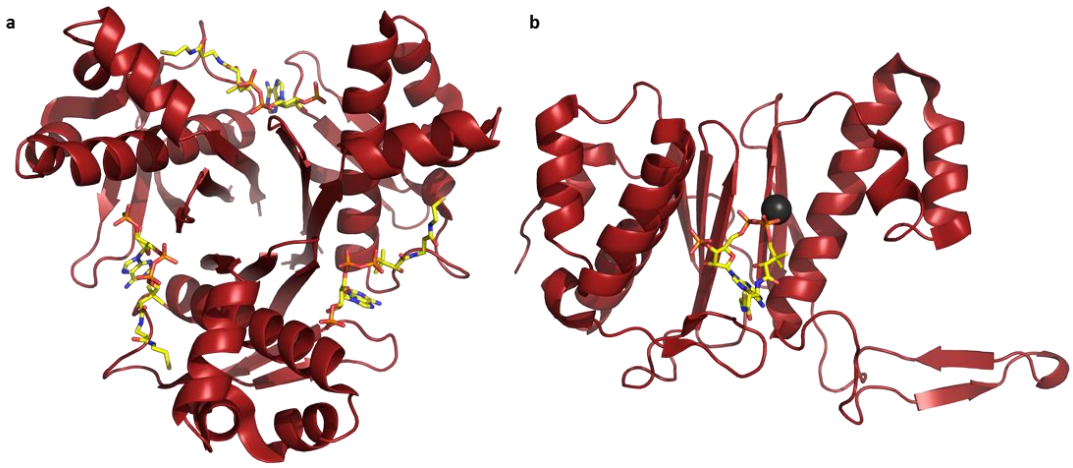


Figure 1.9: Crystal structures of *B. subtilis* (**a**) AcpS, showing the homotrimeric structure with three CoA molecules (yellow sticks) bound at the interfaces between the monomers, PDB code: 1F7L[85] and (**b**) Sfp, showing the pseudodimeric structure resembling two AcpS monomers with a single catalytic site containing CoA (yellow sticks) coordinated by an Mg^{2+} ion (grey sphere), PDB code: 1QR0.[92]

Sfp is a pseudo-homodimer, resembling two monomers of AcpS. However, it only has one active site at the interface of the pseudodimer which is similar to the AcpS active site and also contains a divalent cation. Sfp is much more promiscuous than AcpS and is now used as a tool enzyme for the loading of many different ACPs, both with CoA and other analogues and derivatives as the terminal portion of the CoA molecule is not recognized by this enzyme.[79,83,93–95] For this reason, PPTases in general and especially Sfp are invaluable tools for the study of biosynthetic pathways.

1.4.4 Ketosynthases (KSs)

Ketosynthases are essential components of FAS and PKS biosynthesis which catalyze the formation of new carbon-carbon bonds between two acyl chains. This is a three step reaction beginning with (1) transfer of an acyl chain from an ACP to a conserved cysteine residue in the KS active site, (2) decarboxylation of an ACP bound malonyl group to generate a carbanion and (3) attack of the carbanion on the KS thioester to form a new carbon-carbon bond between the two chains (Figure 1.10).[96] Three conserved active site residues are involved in catalysis: the active site cysteine that retains the acyl chain and two histidine residues that are required for efficient chain translocation and malonyl decarboxylation.[97]

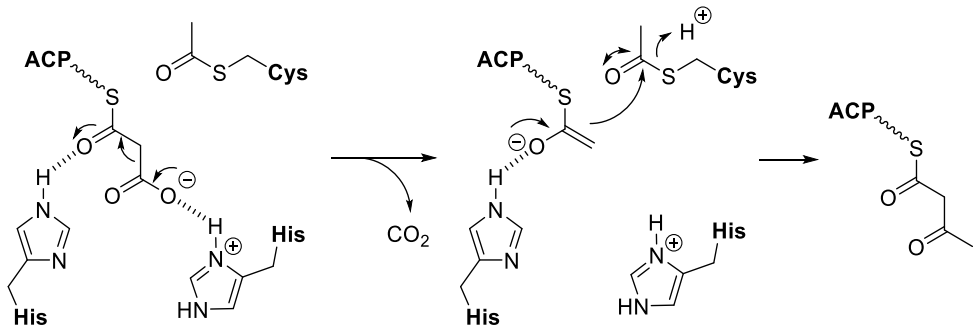


Figure 1.10: The ketosynthase (KS)-catalyzed decarboxylative, carbon-carbon bond forming condensation involving three conserved residues (Cys, His, His) in the KS active site, adapted from Robbins *et al.*[97]

The predicted KS in the tambjamine YP1 pathway is similar to KSs belonging to type II PKS systems. These assemblies contain discrete enzymes with single functions that usually produce aromatic PKs.[98] Type II PKSs also commonly have two different KS domains which have been termed KS_α and KS_β or KS and CLF. The CLF looks very similar to a KS domain but is missing the conserved active site residues required for catalysis. These two domains generally form heterodimers,[99] in contrast to the homodimer KSs of the type II FAS.[63]

The function of the CLF domain is unclear and has been the subject of intensive research. It was designated as a CLF after research into the actinorhodin, granaticin and tetracenomycin KS genes showed that exchange of the CLF from different PKSs resulted in products of altered chain length.[100] However, later research suggested that this wasn't exclusively a function of the CLF but of the whole PKS system.[101] This became clearer with the first and only crystal structure of a KS-CLF heterodimer from actinorhodin biosynthesis (Figure 1.11a).[102] The overall structure of the heterodimer is similar to the FabF dimer of the *E. coli* FAS but is more tightly bound as the two subunits have evolved complimentary contacts. There is only one active site that lies at the dimer interface which creates a tunnel that the growing PK chain extends into (Figure 1.11b). The KS catalytic residues lie at the entrance to the tunnel with the end bound mainly by residues from the CLF. These bulky CLF residues can be mutated to extend the length of the pocket.[103] This results in biosynthesis of longer PKs suggesting that the CLF plays an important role in the control of chain length, utilizing a molecular ruler mechanism.

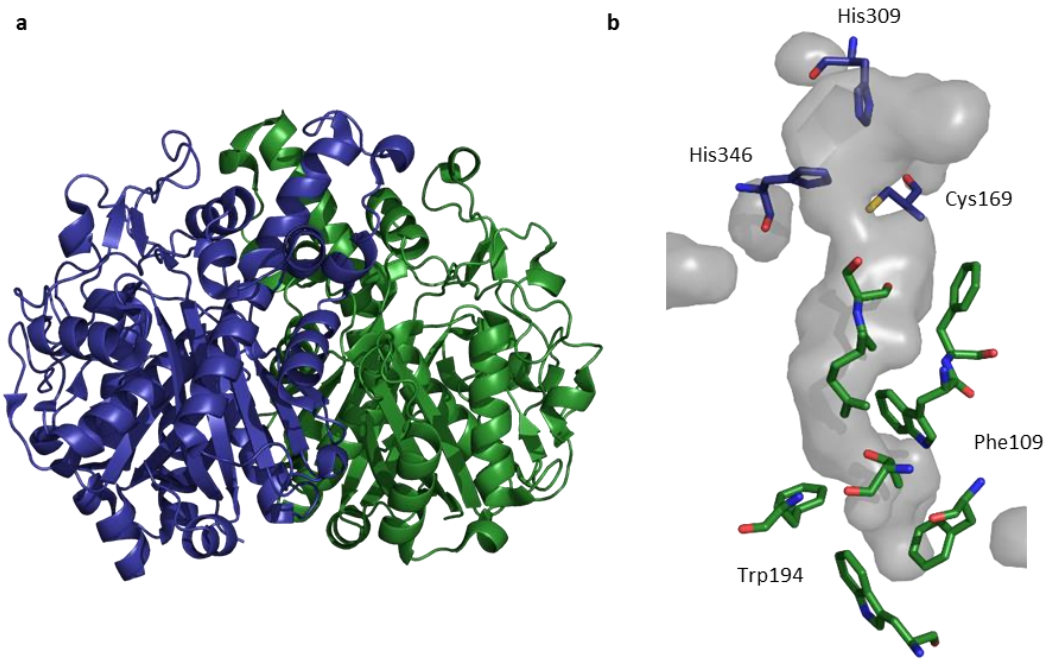


Figure 1.11: Crystal structure of the ketosynthase – chain length factor (KS-CLF) heterodimer from the actinorhodin synthase, PDB code: 1TQY[102] (**a**) overview of the heterodimer with the KS (blue) and the CLF (green) and (**b**) surface representation of the PK binding pocket with the catalytic resides (Cys, His, His) from the KS at the top of the pocket and the bottom bound by residues from the CLF.

Other biochemical assays have proposed that the CLF is also responsible for decarboxylation of malonyl to generate the starter acetyl group for the PKS. As type II PKSs do not have an initiation module, the starter acetyl group needs to be generated from an extender malonyl. In the CLF, the absent KS active site cysteine is often replaced by a glutamine residue which has been shown to promote malonyl decarboxylation.[101] KS proteins have been converted to malonyl-CoA decarboxylases by mutating the active site cysteine to a glutamine or by acetylating the cysteine residue with iodoacetamide.[104,105] This domain has also been hypothesized to have AT activity as it contains the GHS AT motif and type II PKSs usually lack encoded ATs.[106] However, this hypothesis has not yet been explored and many type II PKSs have been shown to rely upon malonyl-CoA:ACP transacylases (MATs) from the native FAS for ACP loading.[107]

1.4.5 Adenylating Enzymes (ANLs)

The first ANL enzyme was identified during exploration of acetate activation in yeast.[108] Berg showed that production of acetyl-CoA proceeded through an adenylate intermediate, as the purified enzyme required acetate for exchange of ATP and pyrophosphate (PP_i) and could turn over synthetic acetyl-adenylate to produce acetyl-CoA. This observation led to the proposal of the now established mechanism of ANL enzymes used to activate carboxylic acids for nucleophilic attack (Figure 1.12).[109]

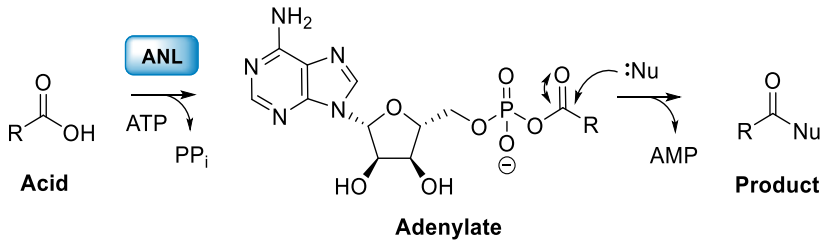


Figure 1.12: Adenylation enzyme (ANL) reaction scheme beginning with attack of the carboxylic acid on the α-phosphate of ATP to form an adenylate intermediate with loss of pyrophosphate (PP_i), then attack of a nucleophile (usually an amine, alcohol or thiol) on the adenylate to form the product amide, ester or thioester with release of AMP.

These enzymes use ATP as the activating agent in two independent half reactions. The first half reaction is attack of the deprotonated carboxylic acid on the α-phosphate of ATP creating an adenylate intermediate; driven by liberation of the PP_i side product. The nucleophile can subsequently attack the adenylate intermediate, releasing adenosine monophosphate (AMP), which is the driving force for the second half reaction. This approach is often used by nature to react weakly nucleophilic carboxylic acids to produce esters, amides, and thioesters. It is commonly used to create ACP and CoA thioesters, which is the predicted method of lauric acid activation in tambjamine YP1 biosynthesis.

There are four identified structural classes of adenylating enzyme. Class I is the largest and best characterized of these families and comprises NRPS adenylation domains, acyl- and aryl-CoA synthetases and luciferases.[109,110] Class II and III are defined by the aminoacyl-tRNA synthetases and adenylation enzymes involved in NRPS-independent siderophore (NIS) synthesis. The newly proposed class IV group of ANLs are the pimeloyl-CoA synthetases (BioWs) involved in the biosynthesis of biotin in certain bacteria.[111,112]

Class I ANLs are those most commonly found in secondary metabolite biosynthesis and several structures in this class have been elucidated. The first structure to be solved was that of firefly luciferase[113] which revealed that ANLs contain two domains, a large N-terminal domain connected to a smaller C-terminal domain *via* a disordered loop. The authors suggested that this protein would have to undergo conformational change to bind the substrate as no binding sites could be observed. These enzymes were also known to undergo drastic conformational changes during catalysis as detected by tritium exchange.[114] The second structure from this class confirmed this hypothesis when an adenylation domain from the gramicidin synthase NRPS was crystallized with a phenylalanine adenylate intermediate bound. A 90° rotation of the two domains with respect to one another was observed, as compared to the luciferase structure.[115] A crystal structure of bacterial acyl-CoA synthetase with a non-hydrolysable adenylate mimic (adenosine-5'-propylphosphate) and CoA showed that the N- and C-terminal domains are rotated ~180° between the adenylation conformation in gramicidin synthase and the observed thiolation conformation. As a result of this data, it was also suggested that the luciferase structure was not functionally relevant and that the adenylation and thiolation conformations are the two catalytic forms.

These two binding modes were further explored when two enzymes, the human medium-chain acyl-CoA synthetase (ACSM2A) and 4-chlorobenzoate CoA ligase were crystallized in both conformations.[116,117] Figure 1.13 shows the structures of ACSM2A first bound to ATP then bound to butyryl-CoA and AMP. The ATP bound structure shows the adenylation conformation with the active site at the interface between the two dimers. The β and γ phosphates are coordinated by an essential Mg^{2+} ion and the yellow P-loop. This loop is conserved in all ATP binding proteins and is responsible for orienting the β and γ phosphates in the active site (conserved motif A3 in NRPS[62] $Z_2X(S/T)(S/T/G)G(S/T)TGXPK$, where Z is a medium hydrophobic amino acid). In ACSM2A, some of these residues can be seen forming hydrogen bonds to the γ phosphate, including the side chains of Thr221, Ser225 and Lys229 and the backbone nitrogen atoms of Gly223 and Ser225 (Figure 1.13b).

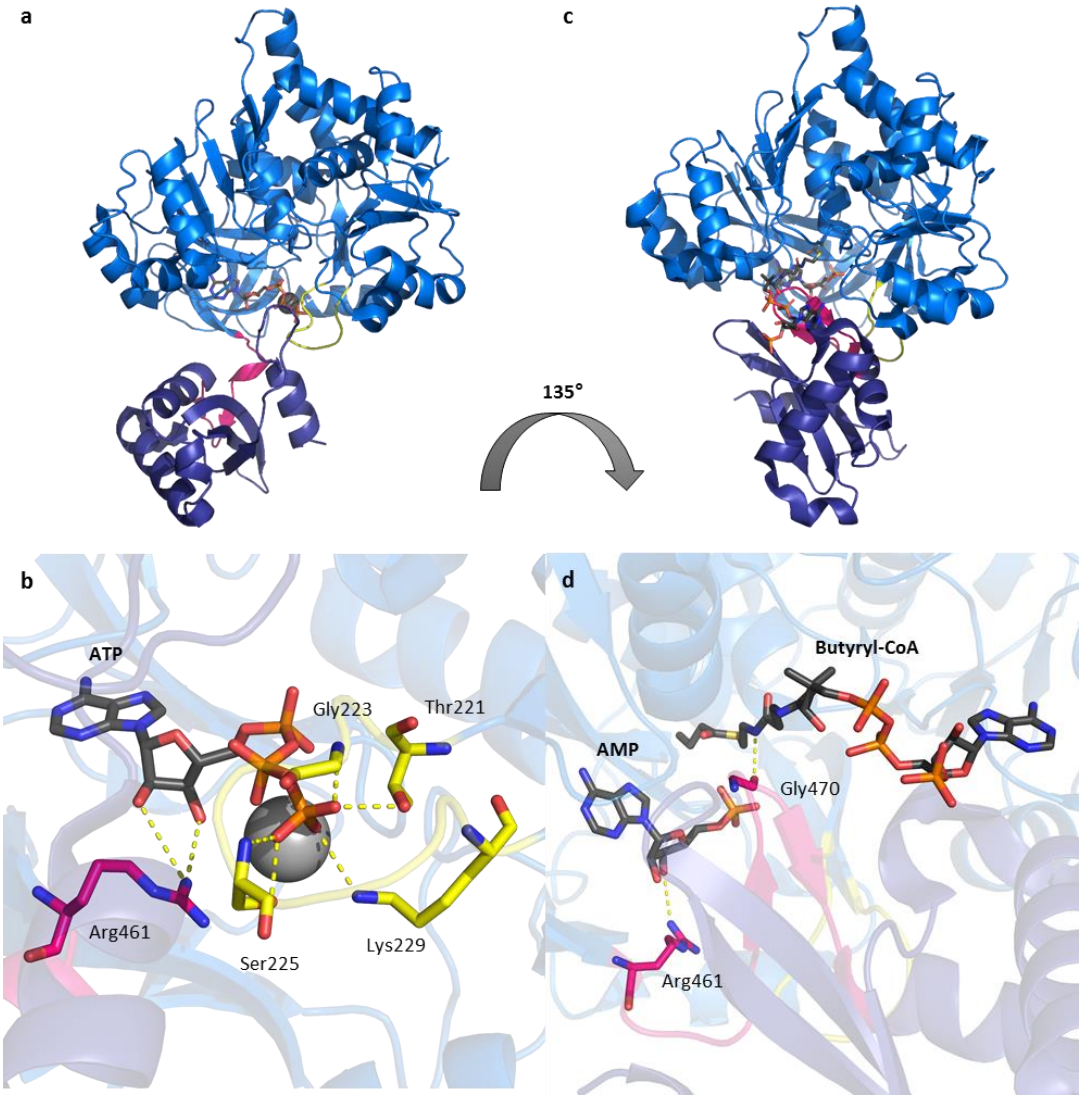


Figure 1.13: Crystal structures of ACSM2A showing the N-terminal domain (blue), the C-terminal domain (purple), the P-loop (yellow) and the hinge region (pink), PDB codes: 3C5E and 3EQ6.[116] (**a**) Full structure with bound ATP (grey sticks) and Mg^{2+} (grey sphere) and (**b**) close up of the active site, (**c**) full structure with AMP and butyryl-CoA (grey sticks) and (**d**) close up of the active site.

Another conserved motif (Motif A8, RX(D/K)X_6G) makes up the hinge region between the two protein domains (shown in pink). This region is flexible and disordered to allow rotation of the domains but contains a conserved arginine residue (Arg461) which interacts with the ribose hydroxyls of ATP. The aspartic acid or lysine residue in this motif is the hinge residue and is essential for conformational change.

The second ACSM2A structure displays the thiolation conformation bound to the reaction products, AMP and butyryl-CoA. Here, the C-terminal domain is rotated 135° with respect to

the N-terminal domain. This results in contraction of the nucleoside binding pocket so that ATP can no longer be accommodated and the P-loop has moved away from the active site. A CoA binding pocket has been created that runs along the interface of the two domains. The conserved glycine residue (Gly470) and other backbone amides of the hinge region hydrogen bond to the amide nitrogen of the 4'-PP. Similar observations, including a 140° rotation of the C-terminal domain were observed for 4-chlorobenzoate-CoA ligase.[117] Mutants of 4-chlorobenzoate-CoA ligase also revealed that areas important for adenylation and thiolation reactions are separate and these residues correlate well with the structural data.[118]

1.4.6 Thioester reductases (TRs)

The end of every natural product assembly line requires a chain release mechanism to liberate the covalently bound product from its attached ACP. The most commonly observed release mechanism for PKS and NRPS natural products is by TE, producing a natural product with a terminal carboxylic acid. However, several other release mechanisms have been identified.[119] One of these is by TR, an enzyme that reduces the ACP-product thioester bond to discharge the extended product as an aldehyde ($2e^-$ reduction) or an alcohol ($4e^-$ reduction).[120] TRs can be NADH[121] or NADPH[122] dependent, using the 'hydride' from this molecule to reduce the thioester. This is the mechanism by which the tambjamine YP1 hydrocarbon tail is proposed to be separated from CoA, producing a fatty aldehyde. This aldehyde is ready for transamination by the downstream TA which is commonly found with TR domains in order to produce amine functional groups.[121]

The first TR domain was discovered by Walsh and coworkers in *Saccharomyces cerevisiae* and is involved in the biosynthesis of lysine.[123] This enzyme releases α-aminoadipate semialdehyde which can subsequently be transaminated to produce lysine in an analogous pathway to tambjamine YP1. Since its discovery several other natural product pathways containing TR domains have been identified producing a wide range of natural products such as the ones identified in Figure 1.14.

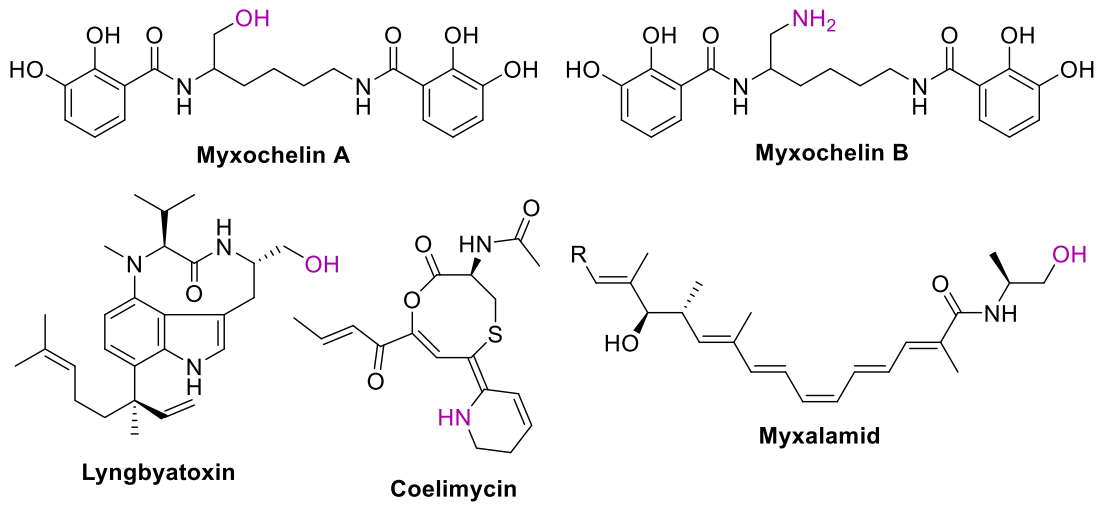


Figure 1.14: Structures of natural products containing functional groups produced by a thioester reductase (TR) or a TR in combination with a transaminase (TA). These functional groups are indicated in purple.

Four structures of TR domains have been solved, one from the *Mycobacterium tuberculosis* NRPS (producing glycopeptolipids),[122] the myxalamid NRPS[124] and two from carboxylic acid reductases (CARs).[125] Each of these structures show the overall fold of the enzyme to be similar to the short chain dehydrogenases (SDRs). The N-terminal NAD(P)H binding region is a Rossman fold containing the GXXGXXG nucleotide binding motif, as can be seen for the *M. tuberculosis* NRPS TR (Figure 1.15a). The small C-terminal region is larger than those of the SDRs and this domain is more variable as it is responsible for binding the CP or CoA substrate. These enzymes also possess the SDR catalytic triad of threonine, lysine and tyrosine (Thr193, Lys232 and Tyr228 in the *M. tuberculosis* NRPS) which have been proposed to directly reduce the CP or CoA bound thioester (Figure 1.15b).[122] Though some have been crystallized as dimers these proteins appear to be monomers in solution, consistent with the SDR structures. All TR structures display a helix-turn-helix (HTH) motif in the C-terminal substrate binding domain which is not present in SDRs. Molecular dynamics of the *M. tuberculosis* TR and its cognate PCP showed that the HTH motif is very dynamic and frequently associates with Helix II (the 'recognition' helix) of the PCP.[124] PCP binding also stabilized the HTH motif and the C-terminal portion of the reductase in general. The 4'-PP arm of the PCP regularly connected with the catalytic triad, lending credence to this binding mode.

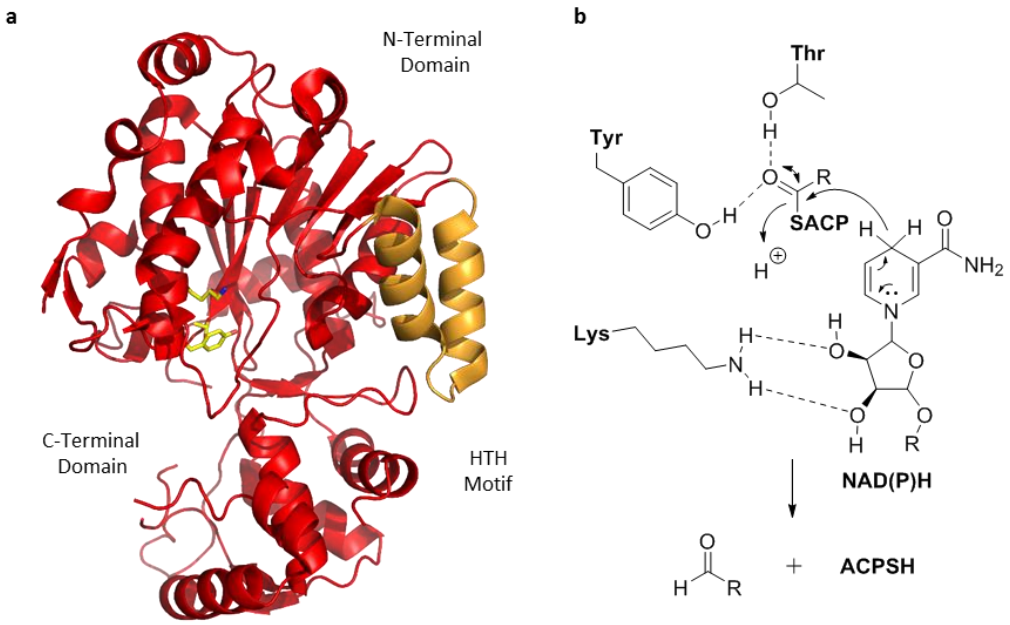


Figure 1.15: (**a**) Crystal structure of a TR domain from *Mycobacterium tuberculosis* NRPS showing the catalytic triad (Thr, Tyr, Lys) in yellow and the helix-turn-helix (HTH) motif in orange, PDB code: 4DQV[122] and (**b**) the proposed mechanism of ACP or CoA thioester reduction adapted from Chhabra *et al.*[122]

Though the TR crystal structures have implied a direct reduction of the thioester bond, biochemical assays with ACP reductases employed in the production of hydrocarbons, have suggested that the acyl chain is initially transferred onto a conserved cysteine residue prior to reduction. This mechanism was proposed by Lin *et al.* who observed a mass increase in *Synechococcus elongatus* ACP reductase after incubation with its cognate acyl-ACP.[126] This hypothesis was supported by the observation that modification with iodoacetamide rendered the enzyme inactive. A similar study was also carried out with the acyl-ACP reductase from *Nostoc punctiforme* which showed the same mass increase.[127] This study also identified the cysteine residue which accepts the acyl chain (Cys294) whose mutagenesis resulted in abolition of enzyme activity. Further research will have to be carried out to establish the exact mechanism of this class of enzyme.

The TR domain from myxochelin (Figure 1.14) biosynthesis has also been extensively studied. Initial assays illustrated that this enzyme could produce not only the aldehyde but also the alcohol form of the product.[128] This demonstrates that the TR can perform a two-step, $4e^-$ reduction from thioester to alcohol which has also been observed in lyngbyatoxin

biosynthesis.[129] This process was later determined to be a non-processive mechanism[122] according to both the enzyme kinetics and crystal structure that indicates the cofactor sits beneath the substrate in the active site. Therefore, the substrate must exit the active site before the cofactor can be recycled. This mechanism is consistent with the observation that several reductases can produce the alcohol product (as in the case of myxalamid[128] and coelimycin[121]) but the aldehyde is the desired product as it is the substrate of a downstream TA. Therefore, the reductase and the TA are in competition for the aldehyde to produce either the alcohol or the amine. This can result in a mixture of these two products, as in the case of myxochelin A and B.[130]

However, biochemical assays of *Mycobacterium marinum* CAR (mmCAR) show that some TRs only produce the aldehyde in a $2e^-$ reduction and are not able to carry out the second step. The Leys group observed a structural reason for this phenomenon as the enzyme was crystallized in two different conformations. In the active conformation, an aspartic acid residue (Asp984) pointed out of the active site, forming contacts with buried residues and positioning NADH next to the Thr and Tyr residues of the catalytic triad. However, this residue can change conformation to point into the NADPH pocket causing surrounding residues to be disordered and the NADPH molecule to move away from the active site. This conformational change is linked to 4'-PP binding and the position of the small C-terminal domain, which is necessary to access the active conformation. This precludes the aldehyde from being turned over as it is no longer associated with the 4'-PP arm. The tambjamine YP1 biosynthetic pathway may use either competition of the transaminase with the TR or indeed this conformational change mechanism to ensure that the desired amine product is produced rather than the alcohol product of the $4e^-$ reduction.

1.5 Pyridoxal 5'-Phosphate (PLP)-Dependent Enzymes

1.5.1 PLP-Dependent Reactions

PLP[131] is a cofactor commonly employed by many different enzymes for the production of both primary and secondary metabolites. PLP exists as a free aldehyde in solution, but binds to enzymes *via* a Schiff base/imine with a conserved lysine residue.[132] This covalent interaction, known as the internal aldimine, anchors the cofactor in the active site in its

resting state (Figure 1.16). It can exist in two tautomeric forms, the enolimine form when the alcohol oxygen is protonated and the ketoenamine form when the imine nitrogen is protonated (Figure 1.16).[133]

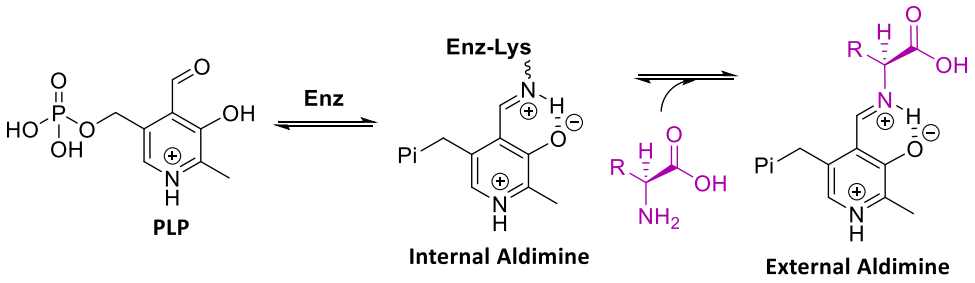


Figure 1.16: The pyridoxal 5'- phosphate (PLP) cofactor and its protein bound, internal aldimine and amino acid bound, external aldimine forms.

PLP dependent enzymes most often use amino acids as substrates and almost all function by a similar mechanism (excluding the PLP-dependent phosphorylases). The amino acid substrate enters the enzyme active site and undergoes transaldimination from the enzyme bound internal aldimine to form a non-covalently bound external aldimine (Figure 1.16). As PLP is a conjugated, positively charged system it has electron sink properties and can easily stabilize a negative charge on the amino acid substrate by forming a quinonoid intermediate. Therefore, this amino acid can undergo many different reactions such as: racemization, α, β or γ elimination or replacement, decarboxylation, aldol cleavage, cyclisation or transamination.[134,135] Dunathan proposed that the reaction specificity depends on the orientation of the substrate, as the bond to be broken must lie perpendicular to the plane of the PLP ring.[136] This conformation results in overlap of the PLP π-orbital with the σ-bonding orbital, allowing the PLP to act as the electron sink when the bond is broken.

Though an abundance of PLP dependent enzymes have been identified and many of their crystal structures solved, they all fit into seven fold types.[132] The vast majority of these are homodimers with some other higher order structures observed. The PLP active site sits at the dimer interface where the cofactor is held in place by a number of hydrogen bonding and electrostatic interactions from both monomers. Most PLP enzymes belong to the fold type I aspartate amino transferase (AAT)[137] family, and these have many different functionalities. Fold types II, III and IV (the tryptophan synthase,[138] D-amino acid aminotransferase[139] and alanine racemase[140] families respectively) also contain several different catalysts, though

these families are much smaller. Finally, fold types V, VI and VIII contain single functionalities including the afore-mentioned PLP-dependent phosphorylases which catalyze reactions through the PLP phosphate group.

AAT was the first enzyme in the fold type I family to have its structure solved.[137] The protein comprises of two domains with the large central domain accommodating a seven stranded β-sheet where catalysis occurs, common to all fold type I enzymes.[132] The type I structures also have a second small domain usually comprising of the C-terminal and sometimes the N-terminal region. However, there is much less homology between the members in the N-terminal section which has led to subgroups being defined within this fold.

1.5.2 α-Oxoamine Synthases (AOSs)

The AOS enzymes catalyze the PLP-dependent reaction between an amino acid and acyl-CoA or ACP (Figure 1.17). This creates a new carbon-carbon bond between the α-carbon of the decarboxylated amino acid and the acyl group, resulting in an α-oxoamine.

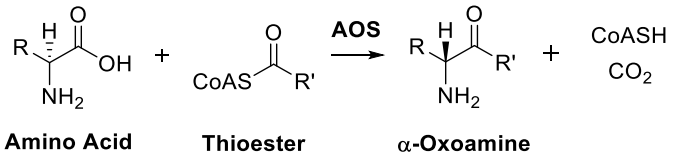


Figure 1.17: General reaction scheme for the α-oxoamine synthases (AOSs) which produce α-oxoamines from amino acids and acyl-CoAs or ACPs.

AOSs are widely found in primary metabolism, contributing to the biosynthesis of heme, biotin (vitamin H) and the membrane components, sphingolipids.[141–144] However, they have also been found in the biosynthetic pathways of secondary metabolites, including the prodiginines, saxitoxin (paralytic shellfish toxin) and fumonisin.[27,145,146] In these pathways the enzymes act as an alternative chain release mechanism to the TEs and TRs, which is the function this enzyme is also expected to play in the tambjamine YP1 pathway.

Biochemical analysis of this family of enzymes has shown a common mechanism for all but one of those studied (Figure 1.18). This is a six step mechanism beginning with (1) transaldimination of the internal aldimine to the external aldimine, (2) abstraction of the

amino acid α-carbon to form a substrate quinonoid, (3) attack of the carbanion on acyl-CoA or ACP, (4) decarboxylation of the amino acid to form the product quinonoid, (5) reprotonation of the product and (6) transaldimination back to the internal aldimine with release of the product.

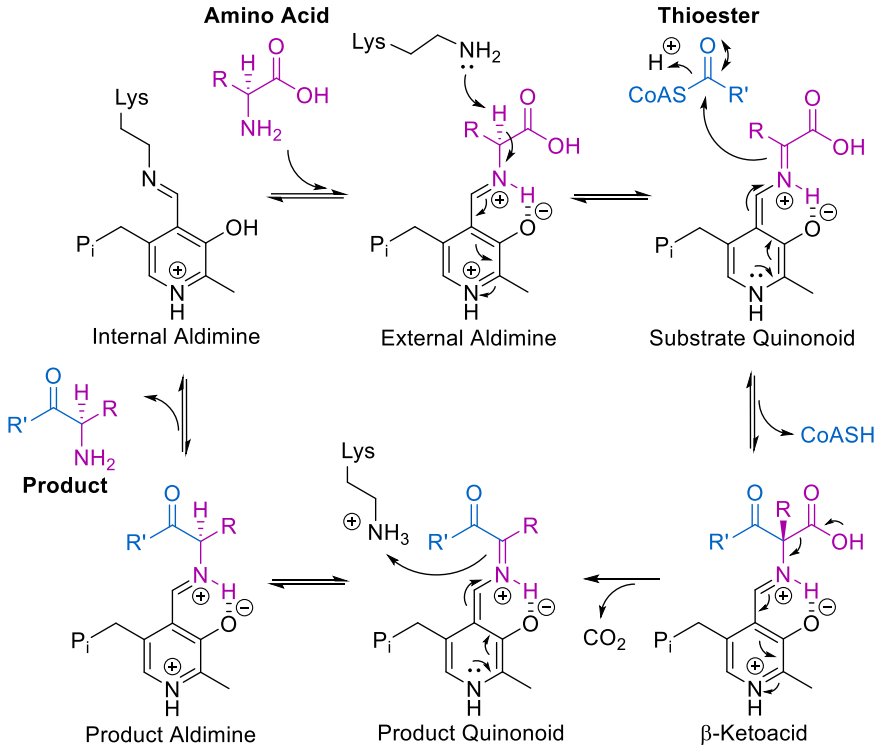


Figure 1.18: Mechanism of α-oxoamine synthase (AOS) reaction beginning with transaldimination, deprotonation of the amino acid, attack on acyl-CoA/ACP, reprotonation of the product and further transaldimination coupled to product release.

This mechanism was elucidated for serine palmitoyl transferase (SPT), the first enzyme in sphingolipid biosynthesis, 5-aminolevulinate synthase (ALAS)[147] involved in heme biosynthesis and 8-amino-7-oxononanoate synthase (AONS),[142,148,149] a biotin biosynthetic enzyme. This reaction is irreversible due to the release of the carbon dioxide molecule. There is only one known exception to this mechanism, which is the reaction catalyzed by ketobutyrate ligase (KBL), a threonine metabolic enzyme.[150] In this pathway, the decarboxylation step is not carried out, instead glycine is condensed with acetyl-CoA to produce a product containing the carboxyl group. Therefore, this is the only enzyme that catalyzes a reversible reaction.

The crystal structures of five of the enzymes from the AOS family have been determined, originating from various organisms (SPT,[93,151,152] ALAS,[153] AONS,[142] KBL[154,155] and CqsA,[156] a biosynthetic enzyme producing a quorum sensing molecule). The crystal structure of *S. paucimobilis* SPT shown in Figure 1.19 illustrates the characteristic, homodimeric type I fold of PLP enzymes. The PLP binding sites located at the interface of the dimer contain several conserved residues that are involved in stabilization of the PLP cofactor.

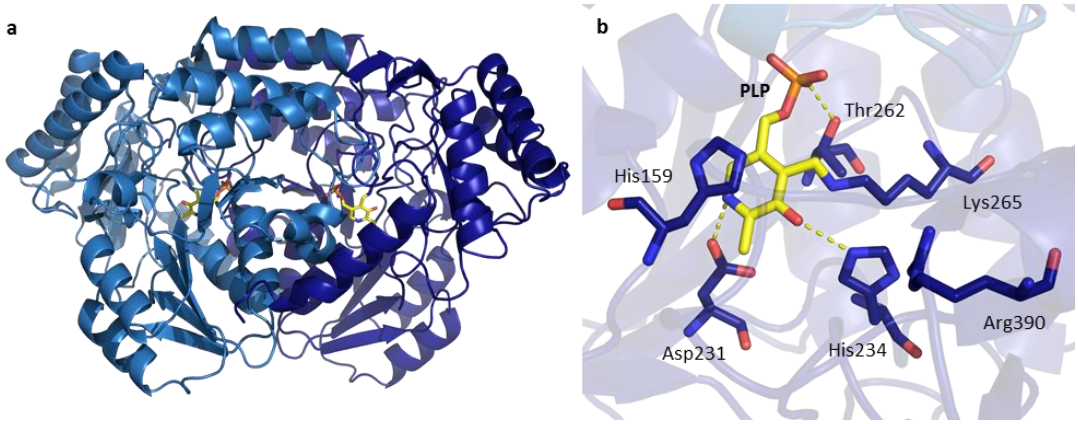


Figure 1.19: Crystal structure of *S. paucimobilis* SPT, PDB code: 2JG2[151] (**a**) overall structure with monomers shown in dark and light blue and (**b**) active site with PLP (yellow sticks) internal aldimine and conserved residues responsible for orienting the cofactor.

The SPT active site shows the universally conserved lysine residue (Lys265) which binds covalently to the PLP cofactor forming the internal aldimine. Asp231 is also conserved across all fold type I enzymes and hydrogen bonds to the pyridine nitrogen. Two histidine residues conserved only in the AOS family are His159 that engages in π-stacking with the PLP ring and His234 that forms a hydrogen bonding interaction with the PLP hydroxyl. Additionally, Thr262 forms a stabilizing hydrogen bond with the PLP phosphate group. Finally, the active site also contains Arg390 which does not appear to form a bond to the PLP while in its resting state but is universally conserved in AOSs. This residue hydrogen bonds to the amino acid substrate carboxylate, orienting the substrate in the active site with the Cα-H bond perpendicular to the PLP plane and is therefore essential for formation of the quinonoid intermediate.[157]

1.5.3 Transaminases (TAs)

TAs employ PLP in a different reaction to catalyze the exchange of an amine group from one molecule to another (Figure 1.20).[133,158] The amine donor is usually an amino acid such as L-glutamic acid or L-alanine that is transformed to α-ketoglutarate or pyruvate, forming pyridoxamine 5'-phosphate (PMP) in the process. The amino acceptor must be a ketone or aldehyde to accept the amino group from PMP, introducing a new stereocenter to the molecule and recycling the cofactor. The transaminases have been exploited for biocatalytic use in several applications as they have excellent enantioselectivity and do not require expensive cofactors.[159]

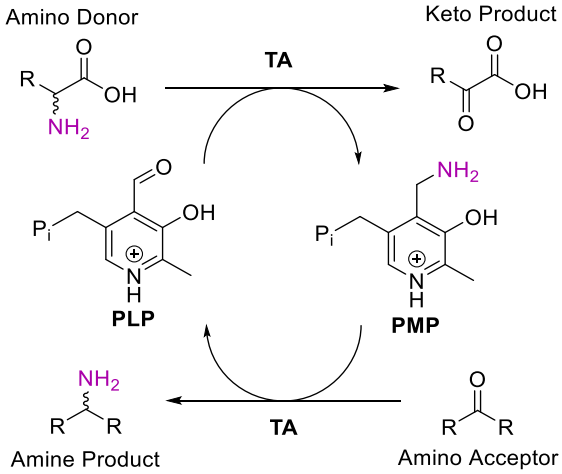


Figure 1.20: Transaminase (TA) reaction scheme beginning with removal of an amine group from the amine donor to produce PMP, then recycling of the cofactor back to PLP by transfer of the amine onto the amino acceptor.

The ping-pong reaction mechanism of transaminases has been well established[160–162] and is shown in Figure 1.21. As with other PLP enzymes, the first step of the mechanism is (1) transaldimination of the lysine-bound internal aldimine to the amino donor-bound external aldimine; (2) abstraction of the α-proton from the amino donor by the displaced lysine residue, forming the substrate quinonoid; (3) transfer of the proton to the PLP carbon to generate a PMP imine and (4) hydrolysis of the imine to yield PMP and the keto coproduct, finishing the first half reaction. The second half reaction begins with (5) condensation of the amino donor with PMP to form an imine; (6) generation of a product quinonoid by the deprotonation of the PMP; (7) protonation of the α-carbon to produce the product aldimine (this occurs stereospecifically on the *re* or *si* face of the PLP[163]) and finally (8) displacement

of the product from the active site by transaldimination with the conserved lysine residue. Each of the steps in this reaction are reversible resulting in an equilibrium between the products and reactants.

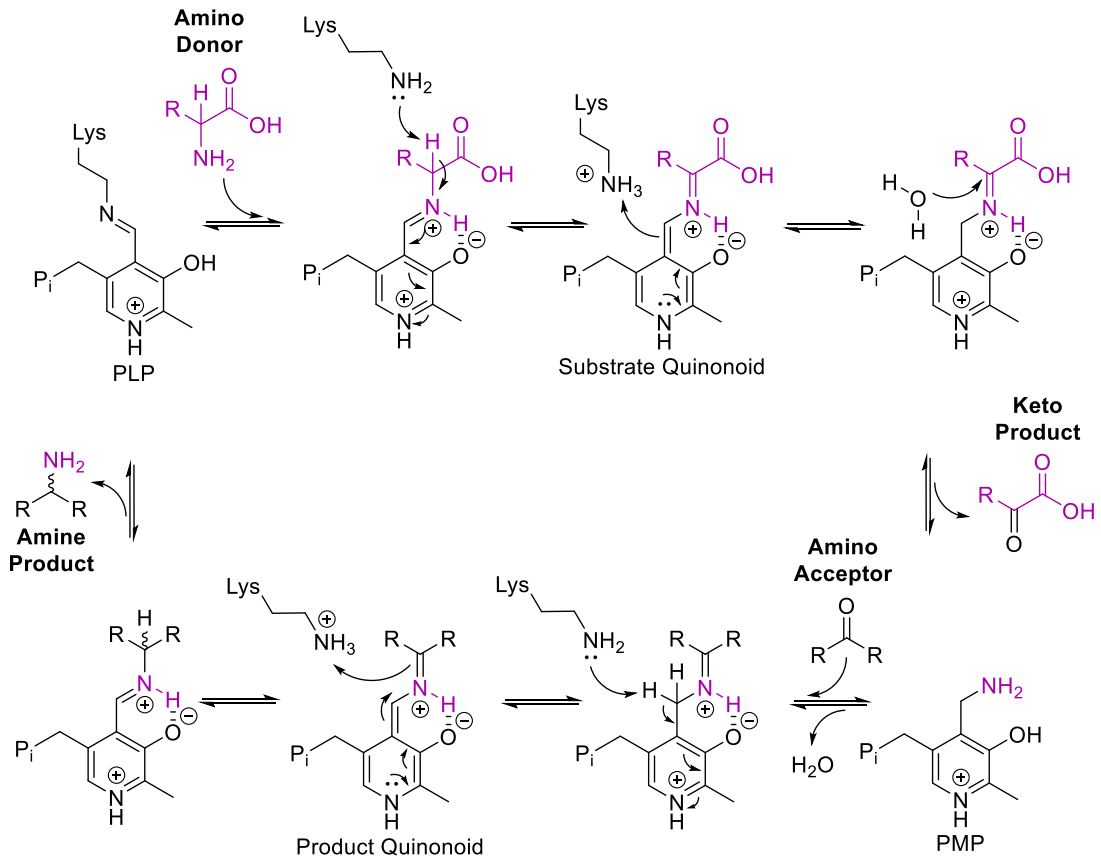


Figure 1.21: Reaction mechanism for transamination by a PLP-dependent transaminase (TA). Steps include, transaldimination, deprotonation of the substrate, protonation of PLP, hydrolysis of the amino donor to yield PMP, condensation of the amino acceptor, deprotonation of PLP, protonation of the product and transaldimination to release the product.

The TAs fall into two of the seven fold types identified for PLP enzymes. These are the fold type I and fold type IV structures. In fold type I, the TAs are grouped into four different subclasses based mainly on their differing N-terminal regions.[158] TamH from the tambjamine YP1 cluster is predicted to be a class III TA. The class III TAs tend to utilize glutamic acid, alanine or more rarely other amino acids as the amine donor.[133] They have dual substrate recognition with amine donors either utilizing different binding pockets or alternative protein conformations to the amino acceptors. This is due to the often large difference in structure between the donor and acceptor substrates.

Ornithine aminotransferase (OAT) was one of the first class III enzymes to be crystallized.[164] Its overall structure is very similar to all fold type I enzymes as it is a homodimer with a large central PLP binding region (Figure 1.22). The active site contains the PLP molecule bound to the conserved lysine residue (Lys292) observed in all PLP-dependent enzymes. It also shows many of the other conserved residues responsible for coordinating PLP in the active site in class III TAs.[165–167]

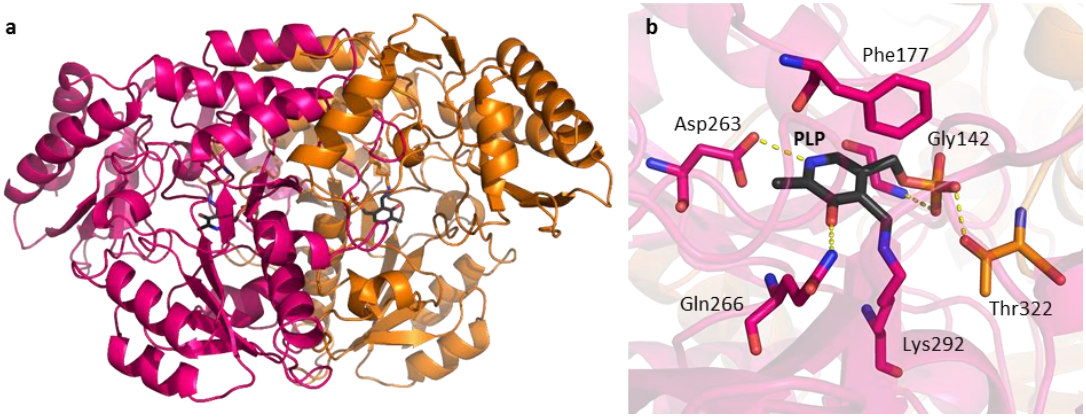


Figure 1.22: Crystal structure of the *H. sapiens* OAT, PDB code: 1OAT[164] (**a**) overall structure with monomers shown in pink and orange and (**b**) active site with PLP (grey sticks) internal aldimine and conserved residues responsible for orienting the cofactor.

The class III TAs also contain the conserved aspartic acid residue (Asp263) that hydrogen bonds to the pyridine nitrogen, as in the AOS SPT structure. The OAT active site includes two amino acids that are conserved among class III TAs that interact with and stabilize the PLP phosphate group. The hydrogen bonds are formed between the phosphate oxygen atoms and the backbone nitrogen of Gly142 and the side chain of Thr322 (deriving from the opposite monomer). In this class, the PLP ring does not form a π-stacking interaction but instead appears to engage in a C-H-π interaction with the usually conserved Phe177 (or other aromatic amino acid) residue. Class III TAs also contain a conserved residue that binds to the PLP hydroxyl group, however, unlike the AOSs this residue is a glutamine (Gln266). Other active site amino acids are responsible for binding and orientating substrates, however, these are highly variable reflecting the wide substrate variability in this class.

1.6 Tambjamine YP1 Summary

The tambjamine YP1 biosynthetic pathway contains at least 11 biosynthetic proteins with various catalytic roles (Figure 1.5). Many of these proteins belong to the previously discussed NRPS and PKS superfamily. However, this pathway also contains alternative chain release and tailoring enzymes such as the AOS TamD and the bispecific TR/TA TamH which are less commonly found in secondary metabolite biosynthesis. None of these enzymes have ever been characterised *in vitro* and many of the transformations postulated in the published biosynthetic pathway have not been observed for tambjamine YP1 biosynthesis or even other similar natural products such as the prodiginines.

Further research into the mechanisms of the enzymes involved in tambjamine YP1 biosynthesis is necessary to fill in the gaps in our knowledge of this biosynthetic pathway. It will also help to gain a better understanding of hybrid biosynthetic systems containing proteins from many and varied families. More in-depth knowledge of the specific enzymes and the substrates that they utilize will also be important in the generation of new biologically active molecules and potentially in the biocatalytic synthesis of many important organic molecules.

2 Aims

The main aim of this work is to characterize the enzymes involved in the biosynthesis of tambjamine YP1 *in vitro*. The biosynthesis of the first bipyrrole intermediate, 4-hydroxy-2,2'-bipyrrole-5-methanol (HBM), has had limited characterization and this intermediate has never been identified. Therefore the first part of this thesis aims to:

- Clone, express and purify the TamD and TamF enzymes involved in the biosynthesis of the second pyrrole ring of HBM
- Characterize these enzymes using mass spectrometry and identify each of the intermediates in the reaction
- Identify the HBM product and monitor its formation
- Probe the substrate specificity of the system for producing alternate bipyrroles and tambjamines

The second part of this work aims to investigate the roles of TamA and TamH in the production of the tambjamine YP1 amine tail by:

- Cloning, expressing and purifying the TamA and TamH enzymes
- Characterizing these enzymes and their products by mass spectrometry
- Confirming the role of the enzymes in the pathway
- Exploring the substrate scope of each of these transformations

3 Biosynthesis of a Bipyrrole

3.1 Introduction

The biosynthesis of 4-hydroxy-2,2-bipyrrole-5-methanol (HBM) in *Pseudoalteromonas tunicata* from its basic building blocks is predicted to require five proteins encoded in the *tam* operon: TamE, TamB, TamG, TamF and TamD.[50] These enzymes utilize L-proline, malonyl-CoA and L-serine to form the first bipyrrole intermediate in the tambjamine YP1 pathway (Figure 3.1).

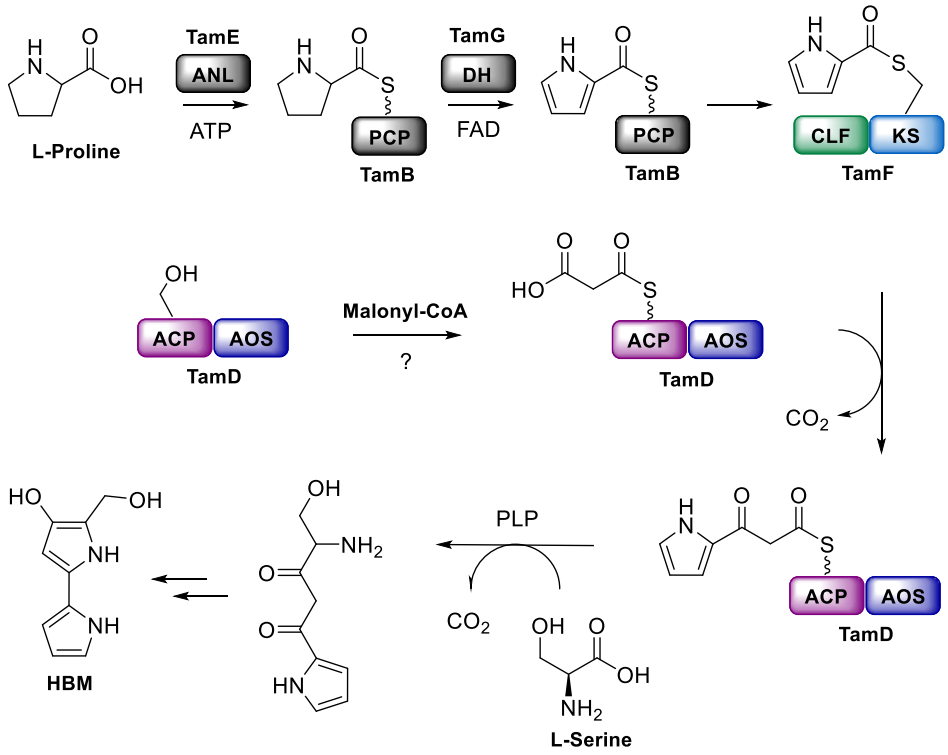


Figure 3.1: Predicted pathway for the biosynthesis of HBM in *P. tunicata* from L-proline, malonyl-CoA and L-serine building blocks by TamE, TamB, TamG, TamF and TamD from the *tam* operon, adapted from Burke *et al*.[50]

In the initial steps, L-proline is activated, then anchored onto the peptidyl carrier protein (PCP, TamB) by an adenylation enzyme (ANL, TamE). It is subsequently oxidized to the pyrrole form by the flavin adenine dinucleotide (FAD)-dependent dehydrogenase TamG, before being passed to TamF, a ketosynthase (KS) – chain length factor (CLF) didomain. Once the pyrrole is bound to the KS active site cysteine, TamF can catalyze the Claisen-like

condensation of malonyl-TamD ACP with the pyrrole. Finally, the TamD α-oxoamine synthase (AOS) uses L-serine and its ACP-bound thioester to catalyze a second Claisen-like condensation and release the extended chain. The terminal amine of this chain can attack the first ketone to cyclise and form HBM in a spontaneous reaction.

Oxidation of a proline ring is a common biosynthetic route to pyrrole in the production of secondary metabolites, so these enzymes have been well characterized in other natural product pathways.[13,55,168] However, the biosynthesis of the second pyrrole ring is an interesting and unique method of creating this moiety. Therefore, the aim of this study was to reconstitute the biosynthesis of the second pyrrole and show that HBM is the first bipyrrole intermediate formed in the tambjamine YP1 and thus also in other 4-methoxy-2,2'-bipyrrole natural product pathways.

To bypass the initial steps of activation and oxidation of proline and focus on the less well characterized transformations carried out by the interesting TamF and TamD didomain proteins, the pyrrole could be fed as a substrate mimic directly to TamF. Many KS enzymes accept *N*-acetylcysteamine (SNAC) thioesters attached to an acyl or peptide chain, a mimic of the terminal portion of 4'-phosphopantetheine (4'-PP).[54,169,170] Utilizing the SNAC substrate mimic, the TamD and TamF reactions could be more easily probed and the biosynthetic pathway to HBM could be reconstituted *in vitro*.

3.2 TamF Protein

3.2.1 Bioinformatic Analysis

Burke *et al*. suggested that the *tamF* gene encodes a KS enzyme that is likely to deliver the first pyrrole ring to malonyl-TamD *via* a Claisen-like condensation, elongating the chain by a two carbon unit.[50] To investigate this hypothesis, the gene sequence was accessed from NCBI (2.3 kb, Appendix 1) and the corresponding protein sequence (Figure 3.2) showed that this gene encodes an 82 kDa protein. The protein sequence was analyzed using the Basic Local Alignment Search Tool (BLAST),[51] which proposed that TamF contains not one but two KS domains, shown on the amino acid sequence in green (residues 1-398) and blue (residues 411-759) respectively. The TamF sequence was also aligned with other enzymes carrying out

the same function in homologous natural product pathways for the biosynthesis of the prodiginines[54,55] (RedX and PigJ, Appendix 2, carried out in ESPript[171]). These sequences show highest similarity in the C-terminal domain where by alignment with other similar known KSs (*Mycobacterium smegmatis* mycocerosic acid synthase (MAS), 5BP1;[172] *Moorea producens* CurL, 4MZ0[173] and *Bacillus subtilis* BacS, 5ENY[174] in Appendix 3) the three conserved residues required for catalysis (Cys507, His637 and His675 in TamF, described in Section 1.4.4) were identified.

```
MSKQKYVISL LDAAIQRGDA ETGMLAVLTD LKQYAKYGIP PIYRNAINRM QLPLLELASE 60
LVTRNKEQLT GRTDVILCAH PGTEQQLQNH YRVTTNAMIR EIMAVTSPST QAQLAAFLPA 120
HSGSSHDKVG EMATTMATRI AQSCQLQGRA FAINSGDNSF AQAISIANDG LKSGKSDAVL 180
VLIANEVLTV SKDTPLAVGA VLLQRSDENH HQDKKAYLHA TQTVSQQGWP ASVDLAAQWY 240
MTSPVNTSQC EPIASSGLIA QPVAEDEQVL GCVAPLAVLL KWLDSDLSSP TMVLSPGQPN 300
EADIALVFGR EPLVFSQAVA PKVVINAQQV WFAGCQGVEA YWQGLNDDQG GMVNIVHEAL 360
ASSQVHVAQG ATFDSYYSNK AALLQPASRD KMGHVAVASV MQTVLESFPT VVLPTNAKGM 420
VITAGNLAPY AQRRVALLPM FTTLTLQIEE VLQANQEVSA QQLLQQWLQQ FAGDAHTKEQ 480
PTWMLSKQIA NFFSKPDWQQ LALEAACAGS IAAIDCAVNA ITSGRVDFAF VAAAEMPVNL 540
HDLCLCSSQQ MLSHSVIATF TEQADGFTPG EGCALILLSR VDATVHLPKL AVIEAIGSST 600
YSKSMIAPNS DGQVNAMRHA FTQTSLLPSD IEFVETHGTG TPIGDLVETQ ALSTVYQASN 660
ERPLNLGALK TQFGHTFAAA GLASVCKVAL CFEHQWQPHN LIRGVLRDQL QLPELNFNPL 720
CQGKPFLSPR GQRHAAVNGF GTGGVNYHLI ISDYCGSQV                       759
```

Figure 3.2: The TamF amino acid sequence (UniProtKB: A4C5W0) encoding a 82 kDa protein with the predicted chain length factor (CLF, green, 43 kDa) and the predicted ketosynthase (KS, blue, 38 kDa) with catalytic residues (Cys507, His637 and His675) in bold.

In contrast, in the N-terminal domains the sequence similarity is low and these sequences also lack the KS catalytic residues, suggesting an alternate role for this domain. Inactive KS domains associated with an active KS often act as CLFs (Section 1.4.4) in type II PKSs that regulate the length of a growing natural product.[103] These domains may also have other roles including malonyl decarboxylation and acyl transfer.[104,106] Therefore, the N-terminal domain of TamF could be a CLF that ensures the extension reaction only occurs once before the product is released by the AOS domain.

3.2.2 Cloning, Purification and Characterization

The *tamF* gene was cloned from *P. tunicata* D2 genomic DNA into a pET22b vector to express the protein with a C-terminal hexahistidine (6xHis) tag (83kDa, recombinant sequence in Appendix 4). Test expression was carried out in BL21 (DE3) *E. coli* cells at 30 °C and 16 °C for 3 h, 5 h and ~18 h at 0.1, 0.5 and 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) concentrations to identify conditions producing the most soluble protein. No soluble protein was observed at 30 °C at any time points, however at 16 °C the soluble protein bands increased in size over the observed time period. This trend may be due to the increased stability of the enzyme at the lower temperatures at which *P. tunicata* grows.[46] Figure 3.3 shows the insoluble and soluble fractions for the 18 h expressions at the various tested temperatures and IPTG concentrations. The IPTG concentration did not appear to affect the amount of soluble protein produced, so gene expression was carried out with 0.1 mM IPTG at 16 °C for ~18 h in 2 L of BL21 (DE3) cells.

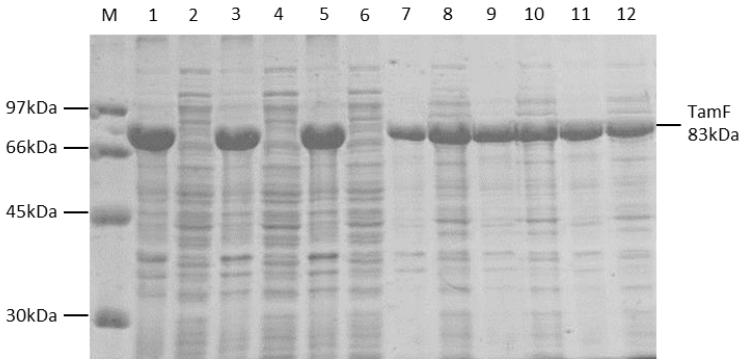


Figure 3.3: SDS-PAGE analysis of the *tamF* pET22b construct ~18 h test expression in BL21 (DE3): (M) Marker (1) insoluble and (2) soluble fractions of expression at 30 °C with 0.1 mM IPTG, (3) insoluble and (4) soluble fractions of expression at 30 °C with 0.5 mM IPTG, (5) insoluble and (6) soluble fractions of expression at 30 °C with 1 mM IPTG, (7) insoluble and (8) soluble fractions of expression at 16 °C with 0.1 mM IPTG, (9) insoluble and (10) soluble fractions of expression at 16 °C with 0.5 mM IPTG and (11) insoluble and (12) soluble fractions of expression at 16 °C with 1 mM IPTG.

The protein was subsequently purified from the cell lysate using nickel immobilized metal affinity chromatography (IMAC) which left the protein ~95% pure but with a small contaminating band (Figure 3.4, lanes 3-8). Therefore, TamF was further purified by gel filtration chromatography (GFC) on a HiLoad 16/60 Superdex S200 column to yield a symmetrical peak at an elution volume of 60.7 mL (Figure 3.4b). Comparison of this elution

volume to a calibration curve of the gel filtration column (Appendix 5) suggests the protein is in a dimeric form. Typical protein yields were 7 mg/L culture after the two purification steps.

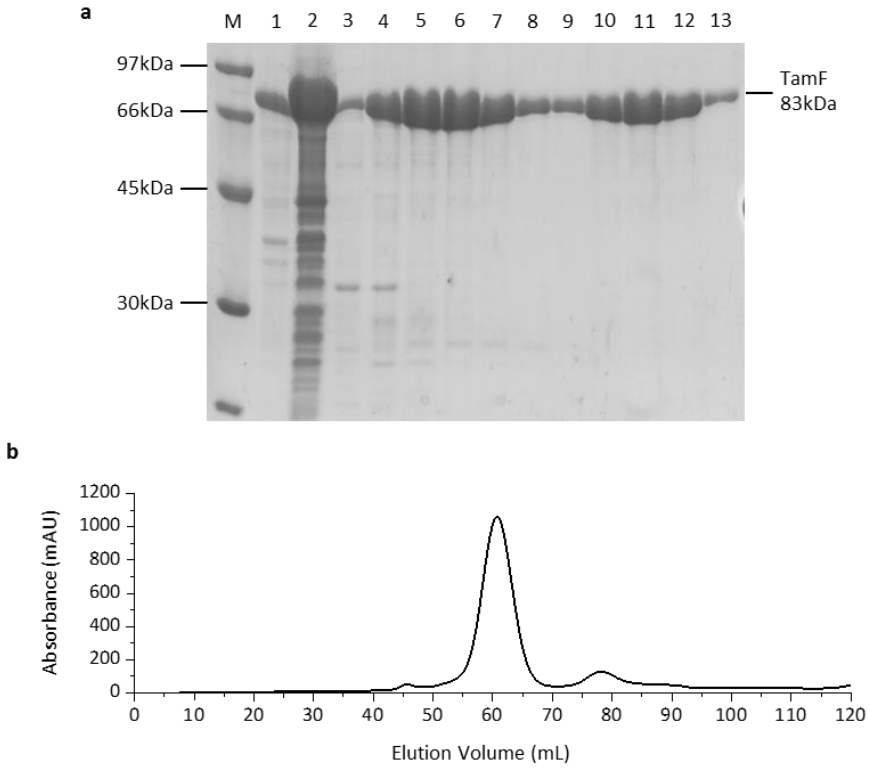


Figure 3.4: TamF purification (**a**) SDS-PAGE analysis: (M) marker, (1) insoluble and (2) soluble fractions of the cell pellet, (3-8) fractions from the nickel IMAC chromatography and (9-13) fractions from the Superdex S200 GFC and (**b**) chromatogram of Superdex S200 GFC showing peak elution at 60.7 mL, indicating that the protein is in dimeric form.

Though FAS KS enzymes are usually homodimers,[175–177] when a KS is associated with a CLF domain as in type II PKSs, the two domains form a heterodimer.[102] In this configuration, catalysis occurs at the dimer interface, allowing the CLF to control the chain length. However, if the protein is already expressed as a fused heterodimer (as with TamF) this would suggest an $\alpha_2\beta_2$ tetramer topology in the globular protein, which is an interesting and unexplored architecture.

The mass of pure TamF was obtained using denaturing liquid chromatography electrospray ionization – mass spectrometry (LC ESI-MS). The resulting spectrum (Figure 3.5) shows several peaks, representing the many charge states of the protein. The molecular mass was

deconvoluted from this spectrum using the maximum entropy (MaxEnt) algorithm from smoothed and centroided data. The observed mass for TamF is 82555.2 Da ±1.6 Da which is consistent with the expected mass of 82554.3 Da calculated from the recombinant protein sequence (benchling.com).

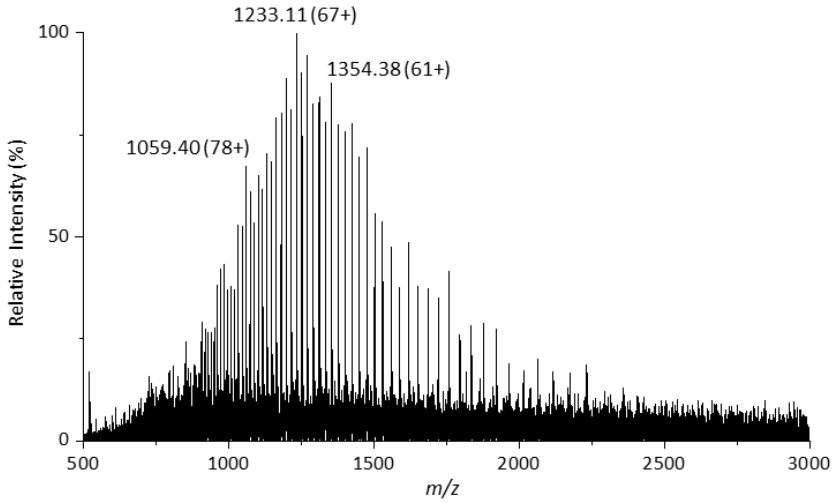


Figure 3.5: Denaturing LC ESI-MS spectrum of recombinant TamF with a deconvoluted mass of 82555.2 Da ± 1.6 Da.

3.2.3 SNAC Substrate Loading

With pure protein in hand, the pyrrole moiety could be loaded onto TamF using a SNAC substrate analogue. This compound was produced by coupling reaction between *N*-acetylcysteamine and pyrrole-2-carboxylic acid using 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC) and 4-dimethylaminopyridine (DMAP, Figure 3.6). The pyrrole SNAC was obtained in 75% yield after purification by flash column chromatography and spectroscopic analysis was consistent with previously reported data.[178]

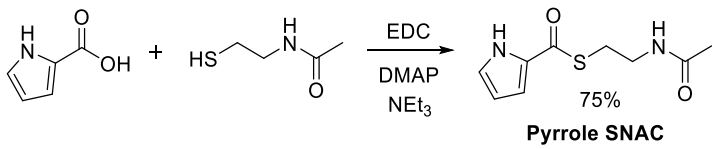


Figure 3.6: Synthesis of pyrrole SNAC by coupling pyrrole-2-carboxylic acid and *N*-acetylcysteamine with EDC, DMAP and triethylamine at room temperature for 16 h.

The resulting pyrrole SNAC thioester (1 mM) was added to TamF (10 µM) before subjecting the protein to LC ESI-MS (Figure 3.7) to determine whether KS loading had occurred.

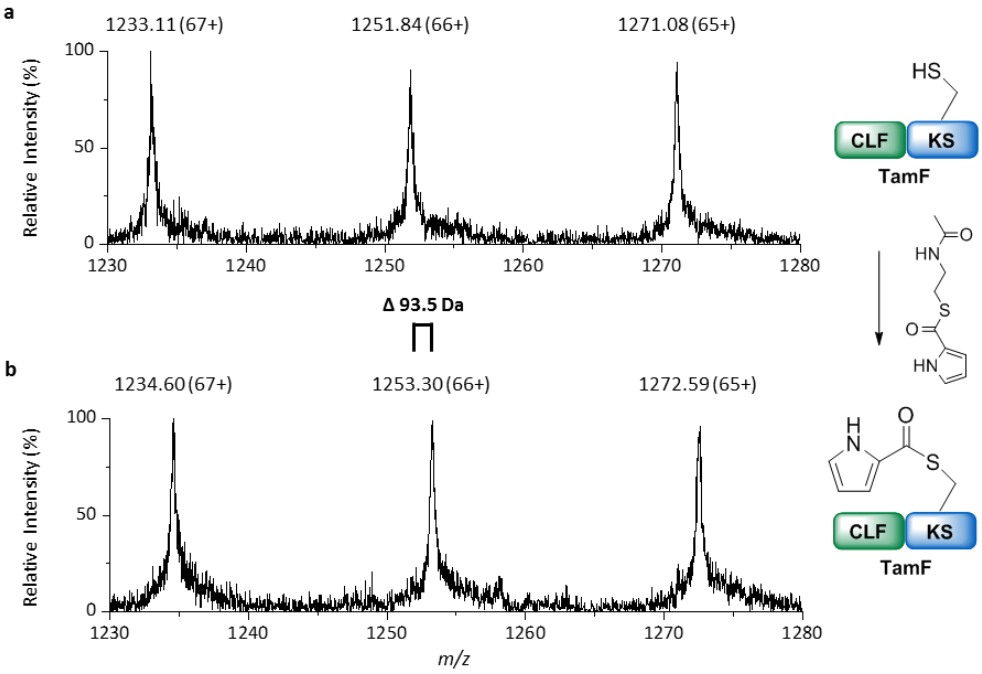


Figure 3.7: Denaturing LC ESI-MS spectrum of recombinant TamF showing the 67 - 65+ charge states (**a**) as purified and (**b**) with 1mM pyrrole SNAC, the mass change is presented as a deconvoluted mass.

After addition of pyrrole SNAC, the TamF peaks are shifted and have a new deconvoluted mass of 82648.7 Da ±1.2 Da. This value is 93.5 Da larger than the unmodified protein, consistent with the predicted mass (93.1 Da) of the addition of a pyrrole group. This data demonstrates that TamF can be easily fully modified with a SNAC thioester 4'-PP mimic in place of the PCP-bound substrate. Using SNAC substrate mimics elegantly avoids the expression and purification of three additional biosynthetic enzymes (TamE, TamB and TamG).

3.3 TamD Protein

3.3.1 Bioinformatic Analysis

Following TamF modification, the KS requires malonyl-TamD to perform chain extension. Bioinformatic analysis of the TamD sequence (1.6 kb gene sequence in Appendix 6, protein

sequence in Figure 3.8) using BLAST indicates that the 58 kDa protein is comprised of two domains: an N-terminal ACP domain and a C-terminal AOS domain, which are shown on the amino acid sequence in purple (residues 1-97) and blue (residues 131-536) respectively. This is unusual within the AOS family as these enzymes tend to have discrete ACP partners or instead utilize CoA substrates.[93,141,179] Between the two functional domains, there is also a linker region whose purpose is unknown and shows no similarity to other protein sequences. Regional order neural network (RONN) analysis (Figure 3.9) predicts a large disordered region of the protein spanning residues 93-129 (highlighted in blue on the TamD sequence). This sequence overlaps with the putative linker region between the two domains. As ACP domains are highly dynamic, this disordered linker could confer the necessary flexibility for the protein to move between the active sites of the upstream TamF enzyme and the TamD AOS domain.

```
MTDNKNTAIE QIHALVIDVV TEQTCYAESD LILDAPMEEG LGIDSIILAS IVSEIQKLFM 60
FETRLNTGSF NTIQALLDIC HNAMLSDAGV QKLAQLGLAA APQAVCVSSQ PEPEQRSTQA 120
QTMRDFVADG SPDLFSKVRK FDQFYKNQAE QGNFWYGMPL SSRCENRATI YDGYQKKERE 180
FLMFASNNYL GLANDPRVIK AICDATQKYG ATNTGCRLIG GTNHLHLELE ARLAAFKGRE 240
ACIVFPSGYS ANLGTISALT GPKDTVISDV YNHMSIQDGC KLSGAKRRIY KHNDMDSLEE 300
VLKGCSESEG GKLIVADGVF SMHGNIVKLP EMVRLARKYQ ARILIDDAHS TGVLGAMGSG 360
TAEHFNLKHE VDLELGTMSK TLAGMGGFVC GDKEVIEYLR FYANSYVFAA TIPANIAAGL 420
IQCIDIIEKE PERISRLRQN ADYLRSALQE CGFNTGDSES AVIPVVIGDE AVAMAMGHQV 480
RQQGMFCQTV VFPGVAVGDA RLRISVLAQH TKEDLDSAIE ILVNSAKTVK LPGFVA     536
```

Figure 3.8: The TamD amino acid sequence (UniProtKB: A4C5W0) encoding a 58 kDa protein with predicted acyl carrier protein (ACP, purple) with its conserved DSX motif (bold) that becomes 4'-PP modified and predicted α-oxoamine synthase (AOS, blue) with conserved lysine residue (bold) that binds PLP. Highlighted in light blue is the regional order neural network (RONN) predicted disordered region.

As with TamF, TamD has homology with enzymes from the analogous natural product pathways of the prodiginines (RedN and PigH, alignment in Appendix 7). However, these enzymes have two ACP domains at their N-termini as opposed to one, both of which can act as carriers for their respective AOS domains.[55] The sequence alignment shows that the AOS domains of these enzymes are very similar, with large areas of identity and similarity. Though the ACP domains have less identity, the TamD ACP is most similar to the second ACP in the RedN and PigH enzymes. The alignment highlights the conserved DSI/V motif (bold, Figure

3.8) which should contain the conserved serine residue that becomes post-translationally modified with 4'-PP.

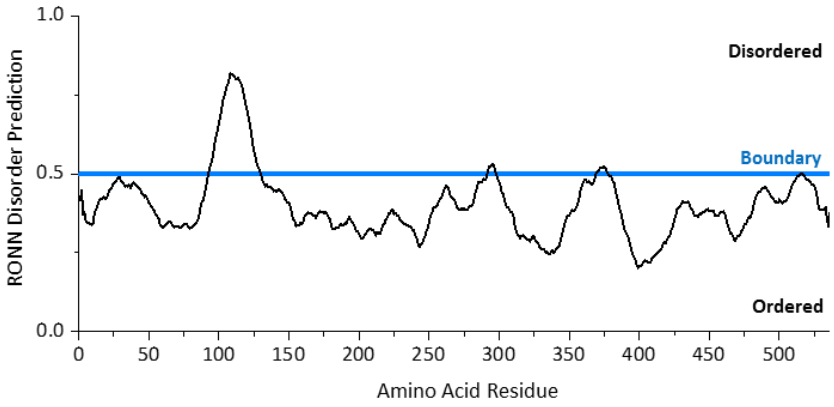


Figure 3.9: Regional order neural network (RONN) analysis of the TamD sequence predicting a disordered region (residues 93-129) overlapping with the predicted linker between the two domains.

Another alignment of TamD with RedN, PigH, *Sphingomonas paucimobilis* SPT, 2JG2,[151] and *E. coli* AONS, 1DJE,[180] (Appendix 8) highlights the many residues conserved within the AOS family (as identified in Section 1.5.2). These include the conserved lysine residue that binds PLP (Lys380, bold in Figure 3.8), several other residues involved in hydrogen bonding to the PLP ring and Arg503 which is essential for catalysis and formation of the quinonoid intermediate.

3.3.2 Cloning, Purification and Characterization

A *tamD* construct in pEHISTEV,[181] previously prepared by Daynea Wallock-Richards was used to produce recombinant TamD with a Tobacco Etch Virus (TEV)-cleavable 6xHis tag (Appendix 9). The gene was expressed with 0.1 mM IPTG at 20 °C for 4 h in 2 L of BL21 (DE3) cells. The protein was purified from the cell lysate (Figure 3.10) using nickel IMAC before subjecting it to cleavage by TEV protease. The TEV protease cleavage removed the 6xHis tag and a second nickel IMAC purification allowed cleaved TamD to be collected in the flow-through separated from the 6xHis tag, uncleaved TamD and TEV protease bound to the column. Finally, GFC purification provided a symmetrical peak at 64.0 mL with an excellent yield of 22 mg/L culture. The elution volume is that expected of the protein dimer, consistent with AOS and other PLP-enzymes which are almost all dimers.[135]

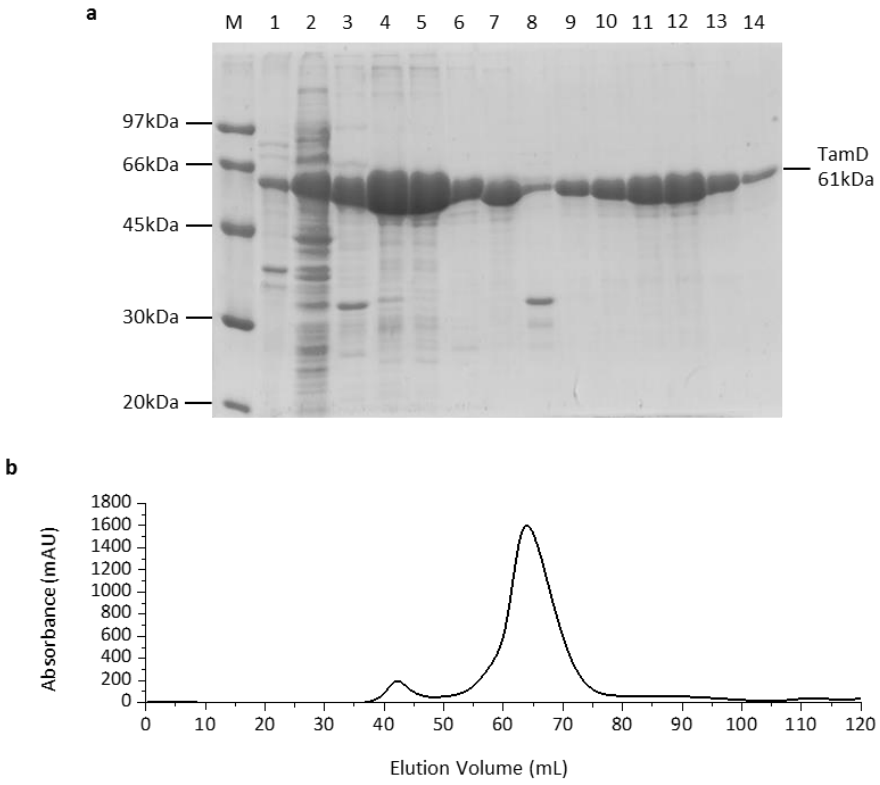


Figure 3.10: TamD purification (**a**) SDS-PAGE analysis: (M) marker, (1) insoluble and (2) soluble fractions of the cell pellet, (3-6) fractions from the nickel IMAC chromatography, (7) wash of TEV-cleaved TamD, (8) elution of uncleaved TamD and TEV protease (9-14) fractions from the Superdex S200 GFC, (**b**) chromatogram of Superdex S200 GFC showing peak elution at 64.0 mL, elution volume of a protein dimer.

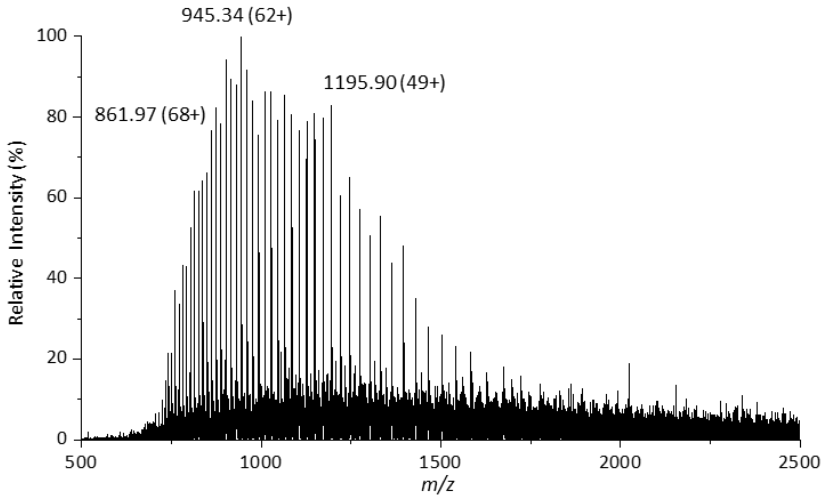


Figure 3.11: Denaturing ESI-MS spectrum of recombinant TamD with a deconvoluted mass of 58550.4 Da ± 0.4 Da.

LC ESI-MS analysis of TamD showed the deconvoluted mass expected of the *apo*-protein at 58550.4 Da ± 0.4 Da consistent with the calculated mass of 58550.3 Da (sequence in Appendix 9). This confirmed that the endogenous *E. coli* phosphopantetheinyltransferase (PPTase) is unable to modify TamD with 4'-PP during expression since this would have added a mass of 340 Da.

3.3.3 Post Translational Modification

As the protein was expressed in its unmodified form, the ACP domain requires post-translational modification (PTM) with 4'-PP from CoA. In *P. tunicata* this PTM is likely to be carried out by a PPTase encoded in the *tam* operon (TamS). This enzyme was cloned in four different constructs with N- and C-terminal 6xHis tags, with no tag and as a maltose binding protein (MBP) fusion. Only the MBP fusion produced soluble protein, however, when the protein was purified it began to rapidly degrade and was unable to modify TamD.

Instead, the well-known promiscuous PPTase Sfp from *B. subtilis* was employed (Section 1.4.3).[182] The Sfp reaction was carried out with CoA and Mg^{2+} at 4 °C for ~18 h which resulted in a TamD deconvoluted mass shift from 58550.4 Da ± 0.4 Da to 58890.9 ± 1.1 Da (Figure 3.12a,b). This is an increase of 340.5 Da, consistent with the mass of 4'-PP (340.1 Da), indicating complete modification of TamD from its *apo*- to its *holo*-form.

The TamD ACP also requires a malonyl unit bound to the 4'-PP before reaction with TamF. It is unclear where the malonyl group originates from in the native organism and how it is transferred onto the ACP. As some ACPs are able to self-acylate from a CoA derivative,[183] *holo*-TamD was incubated for ~18 h at 4 °C with 500 µM malonyl-CoA to determine if this ACP is able to self-acylate. There was a change observed in the spectrum (Figure 3.12c) as a peak appeared to the right of *holo*-TamD with a deconvoluted mass increase of ~87 Da. This is possibly malonyl-TamD (predicted mass Δ86 Da) but since the mass difference is small and the protein is large and highly charged, the changes are difficult to confirm with absolute accuracy.

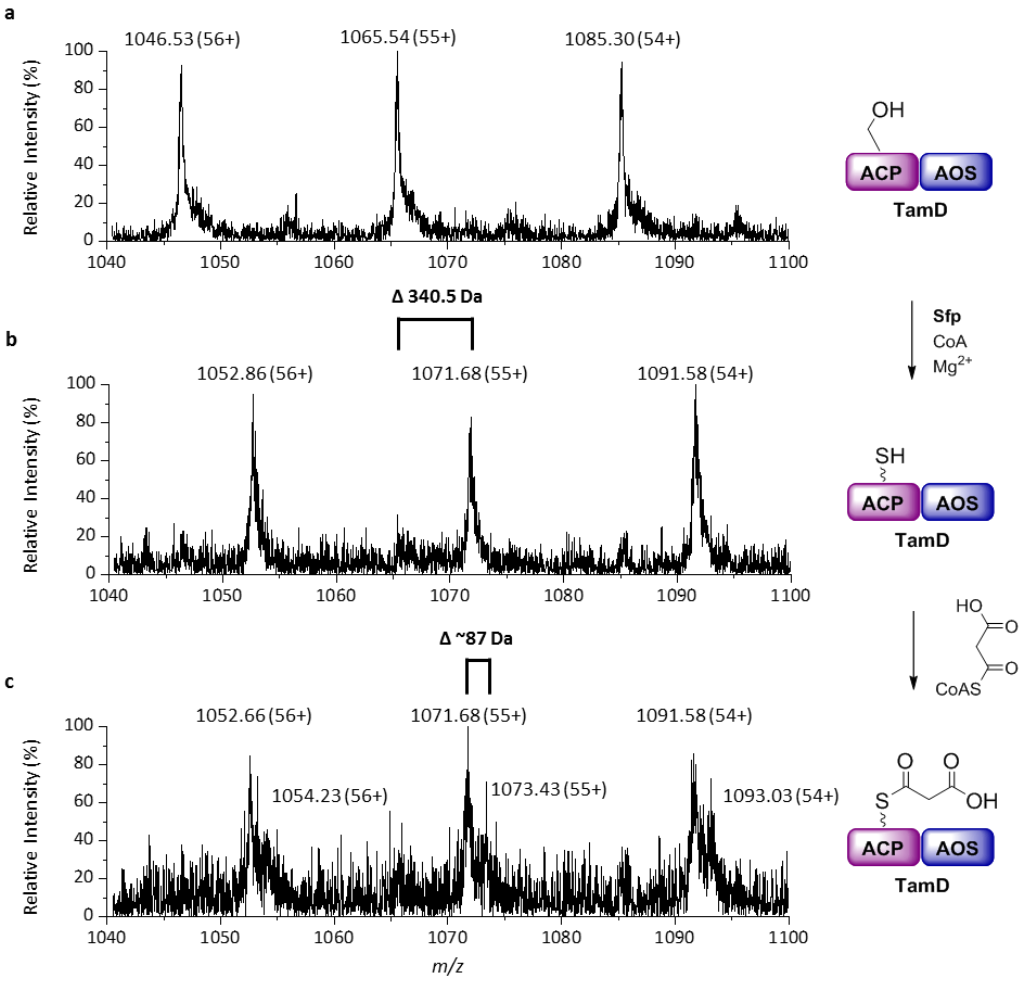


Figure 3.12: Denaturing ESI-MS spectrum of recombinant TamD showing the 56 - 54+ charge states (**a**) as purified, (**b**) after incubation with Sfp, CoA and Mg^{2+} at 4 °C for ~18 h and (**c**) after further incubation with malonyl-CoA at 4 °C for ~18 h.

3.3.4 ACP Cloning, Purification and Characterization

As ACP domains are small at around 10 kDa, it is much easier to track covalently bound modifications in these proteins as they have fewer charge states by ESI. A clone provided by Daynea-Wallock Richards of the TamD ACP domain included TamD residues 1-95 without an affinity tag. However, this domain was difficult to purify and lacked stability, as it would begin to precipitate after purification. Without a 3D structure of TamD, it is unclear where the domain boundaries lie, however, BLAST analysis suggested the ACP domain spans residues 1-97 so extra residues may be necessary for protein stability.

Therefore, new TamD ACP clones of different lengths (with 3, 6 and 9 amino acid extensions, sequences in Appendix 10) were produced with TEV cleavable 6xHis tags: ACP_{95}, ACP_{98}, ACP_{101} and ACP_{104} to determine if increasing chain length improved stability. Test expression of the new constructs (Figure 3.13) showed clearly that the ACP domain requires more than 95 residues to produce a stable protein. No soluble protein can be observed for ACP_{95} whereas the majority is soluble for ACP_{98} and above. There was no significant difference in solubility after increasing the ACP length beyond 98 residues, suggesting the domain boundary is consistent with the BLAST prediction.

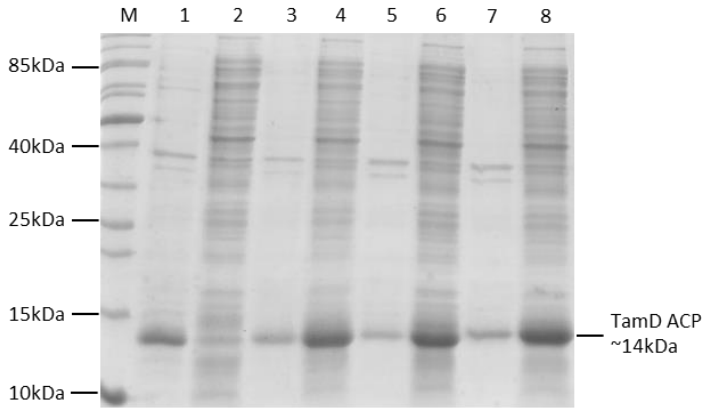


Figure 3.13: SDS-PAGE analysis of test expression of TamD ACP constructs in BL21 (DE3): (M) Marker (1) insoluble and (2) soluble fractions of expression of TamD ACP_{95}, (3) insoluble and (4) soluble fractions of expression of TamD ACP_{98}, (5) insoluble and (6) soluble fractions of expression of TamD ACP_{101}, (7) insoluble and (8) soluble fractions of expression of TamD ACP_{104}.

Therefore, the TamD ACP_{98} construct was used for further experiments and was expressed with 0.1 mM IPTG at 16 °C for ~18 h in 2 L of BL21 (DE3) cells. The ACP_{98} protein was purified by the same method as TamD: using nickel IMAC, TEV cleavage, a second nickel IMAC and finally GFC (Figure 3.14). GFC for this protein was carried out on a HiLoad 16/60 Superdex S75 column for low molecular weight proteins. TamD ACP_{98} GFC produced a peak at 72.7 mL, an elution volume expected of the protein monomer (calibration curve in Appendix 11). A monomeric protein is to be expected for a stand-alone ACP as the TamD protein is likely to dimerise *via* the AOS domain.[135] The purification yielded only 2 mg/L culture of the TamD ACP however, this was sufficient for further analyses due to its small molecular mass (11 kDa).

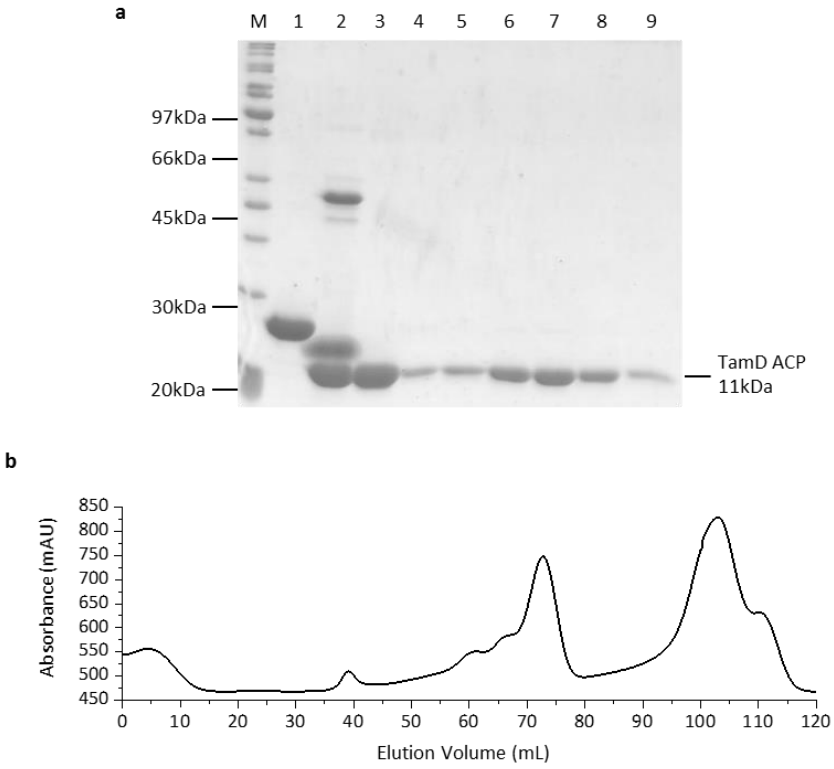


Figure 3.14: TamD ACP_{98} purification (**a**) SDS-PAGE analysis: (M) marker, (1) protein after nickel IMAC purification, (2) after TEV protease cleavage, (3) wash of TEV protease cleaved TamD ACP_{98}, (4-9) 3 mL fractions from the Superdex S75 GFC, (**b**) chromatogram of Superdex S75 GFC showing peak elution at 72.7 mL, indicative of a protein monomer.

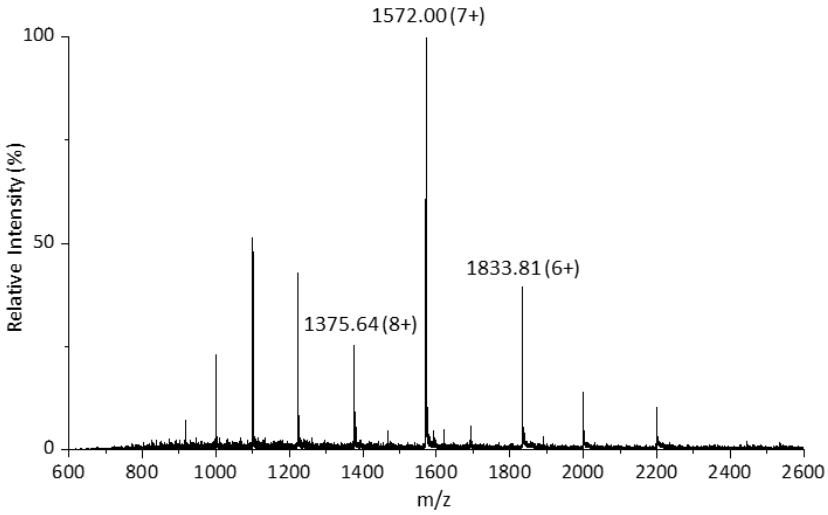


Figure 3.15: Denaturing ESI-MS spectrum of recombinant *apo*-TamD ACP_{98} domain with a deconvoluted mass of 10997.0 Da ± 0.1 Da.

TamD ACP_{98} was also subjected to LC ESI-MS analysis and produced a spectrum (Figure 3.15) with a deconvoluted mass of 10997.0 Da ± 0.1 Da, which is consistent with the predicted *apo*-protein mass of 10997.5 Da. The protein mass spectrum now shows far fewer charge states which are better separated compared to the full-length TamD (Figure 3.11).

3.3.5 ACP_{98} Post Translational Modification

The TamD ACP_{98} could now be post-translationally modified with *B. subtilis* Sfp. The ACP domain was subjected to the same reaction containing Sfp, CoA and Mg^{2+} as the full-length enzyme with the equivalent result of complete modification to the *holo*-form. To speed up the modification process and to avoid having to express and purify another protein, TamD ACP_{98} was subsequently coexpressed with untagged Sfp protein. Coexpression resulted in an increase of the deconvoluted mass of TamD ACP_{98} to 11336.6 Da ± 0.4 Da (Δ 339.6 Da) consistent with the 340.5 Da increase expected of 4'-PP modification and *apo*-ACP_{98} is no longer observed (Figure 3.16).

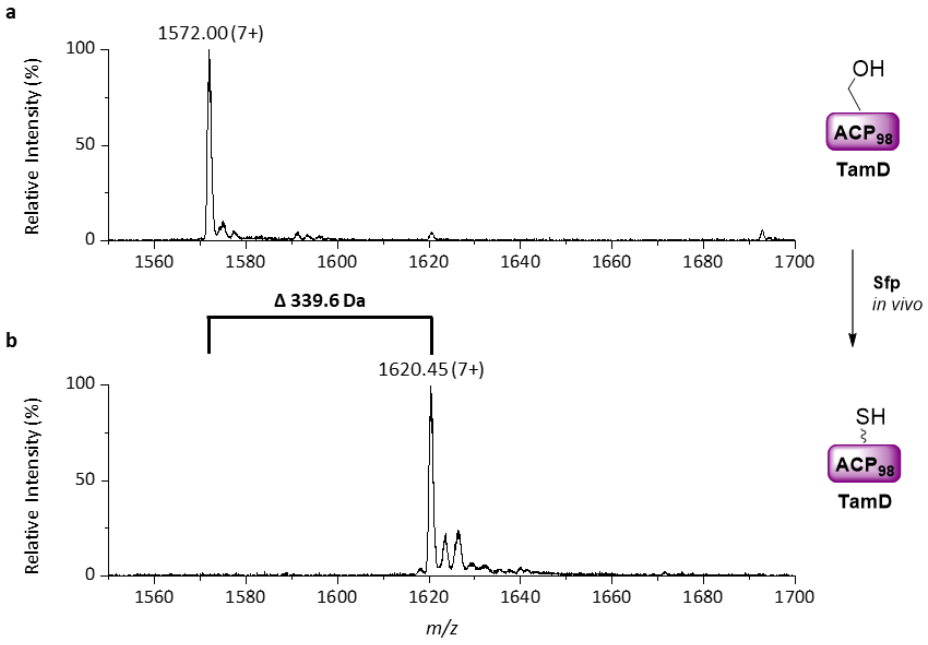


Figure 3.16: Denaturing ESI-MS spectrum of the recombinant TamD ACP_{98} domain showing the 7+ charge states (**a**) as purified and (**b**) after coexpression with Sfp, the mass change presented is a deconvoluted mass.

3.4 TamF and TamD ACP$_{98}$ Claisen-like Condensation

Previously, loading of TamF with its pyrrole substrate was achieved with pyrrole SNAC (reaction **A**) and *holo*-TamD ACP$_{98}$ was produced *in vivo*. However, TamD ACP$_{98}$ also needs to be loaded with a malonyl group (reaction **B**) before Claisen-like condensation can occur between these two enzymes (reaction **C**, Figure 3.17). The malonyl loading reaction was previously attempted with full-length TamD by self-acylation (Figure 3.12). The reaction appeared to be very slow, as TamD was allowed to self-acylate for ~18 h and reached less than 50% conversion. However, many studies of type II PKSs have shown that they are reliant on malonyl-CoA:ACP transacylases (MATs) from their endogenous fatty acid synthase (FAS).[99,102,184] These enzymes transfer the malonyl group from malonyl-CoA to ACPs, usually for fatty acid chain extension, but can also be employed in the type II PKS. Though *P. tunicata* is likely to have its own MAT enzyme, FabD, the *E. coli* MAT was available in the laboratory (provided by Alex Ekström). Therefore, this enzyme was employed to attempt catalysis of acyl transfer from malonyl-CoA to TamD ACP$_{98}$ (reaction **B**, Figure 3.17).

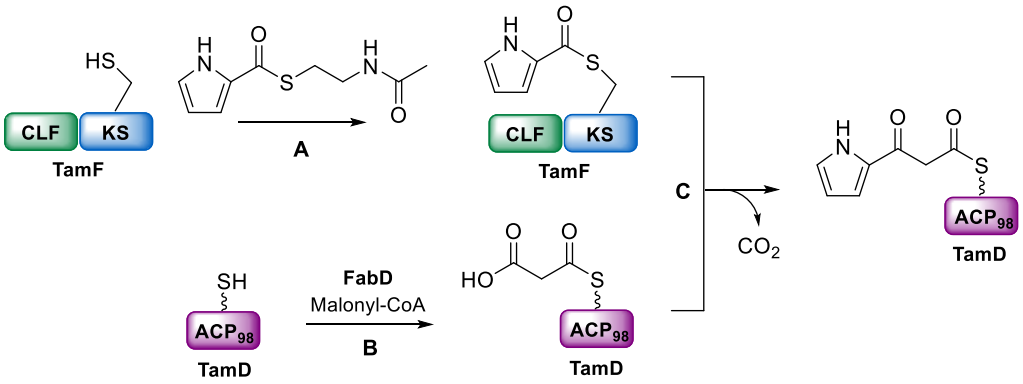


Figure 3.17: Reaction scheme for the loading of TamF with pyrrole SNAC (**A**), TamD ACP$_{98}$ with malonyl-CoA catalyzed by *E. coli* FabD (**B**) and the Claisen-like condensation between these Tam proteins to extend the chain (**C**).

When *holo*-TamD ACP$_{98}$ is incubated with only malonyl-CoA for 15 min, a very small malonyl peak appears at 11423.6 ± 0.1 Da (pink, Figure 3.18b) as compared to the control (Figure 3.18a). Excitingly, with addition of *E. coli* FabD, around 50% of the ACP is converted to the malonyl form after the same period of time (Figure 3.18c). This demonstrates that the TamD ACP$_{98}$ domain is a substrate for FabD and can be used to speed up the acylation reaction.

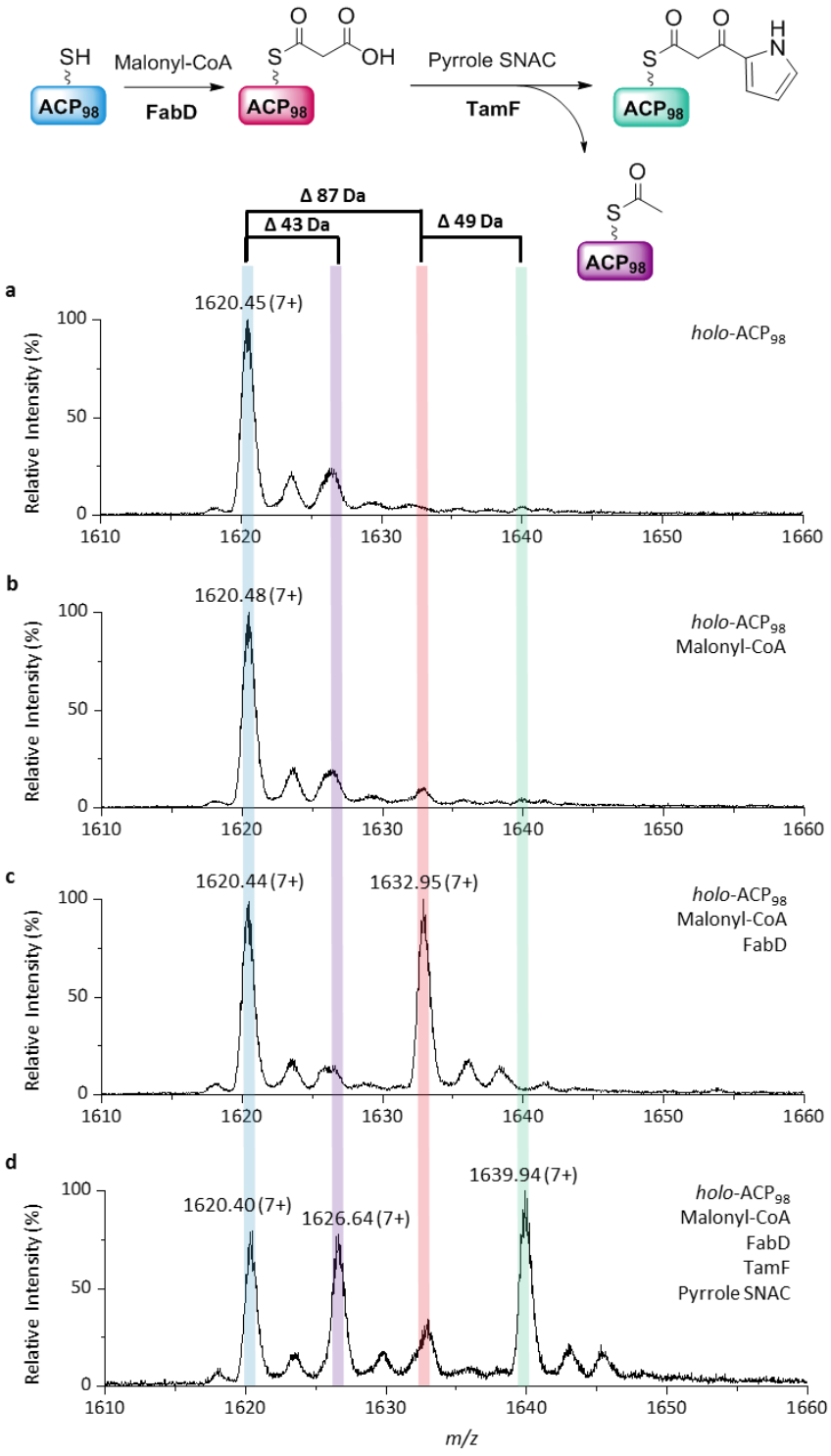


Figure 3.18: Denaturing ESI-MS spectrum showing the 7+ charge state of (**a**) *holo*-TamD ACP_{98}, (**b**) *holo*-TamD ACP_{98} after 15 min incubation with malonyl-CoA, (**c**) *holo*-TamD ACP_{98} after 15 min incubation with malonyl-CoA and *E. coli* FabD and (**d**) *holo*-TamD ACP_{98} after 15 min incubation with malonyl-CoA, *E. coli* FabD, TamF and pyrrole SNAC.

Once TamD ACP_{98} malonation was achieved, TamF and the ACP could be coupled in the fully reconstituted chain extension reaction. Addition of TamF and pyrrole SNAC to the reaction along with malonyl-CoA and *E. coli* FabD resulted in the appearance of a new peak (green, Figure 3.18d), with a deconvoluted mass 49 Da larger than the malonyl peak. This mass is consistent with malonyl decarboxylation and addition of the pyrrole ring from TamF, demonstrating that the system is functional and the first Claisen-like condensation between Tam enzymes has been achieved. Additionally, a second peak also appears (purple) which is smaller in mass than the malonyl peak. This deconvoluted mass is 43 Da larger than *holo*-ACP_{98} corresponding to acetyl-ACP (predicted mass: Δ 42 Da) which would be produced if TamF catalyzed an unproductive decarboxylation. The appearance of acetyl-ACP was unexpected because it suggests the reaction is not particularly efficient, with acetyl-ACP accounting for nearly 50% of the product. The acetyl-ACP product is not observed when the TamF reaction is carried out in the absence of pyrrole SNAC, illustrating that the KS must be acylated before it catalyzes decarboxylation.[105] However, since the product was observed, the Claisen-like condensation was successful on the stand-alone ACP domain. Though this could not be monitored on full-length TamD, the assumption was made that this would also occur on the didomain protein.

3.5 TamD Amino Acid Binding

Once the TamD ACP is primed with the correct substrate, the AOS domain must bind L-serine before carrying out a Claisen-like condensation between the amino acid and the ACP-bound thioester (Figure 3.1). Binding of amino acid substrates to pyridoxal 5'-phosphate (PLP)-dependent enzymes can be monitored through the ultraviolet (UV)/visible absorption profile of the PLP cofactor. PLP exists in two isoforms when bound as a Schiff base: the enolimine and ketoenamine forms (Figure 3.19).[93]

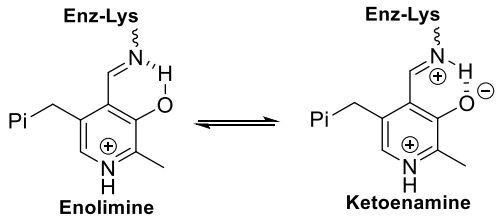


Figure 3.19: Enolimine and ketoenamine forms of enzyme-pyridoxal 5'-phosphate (PLP) Schiff bases.

The enolimine form absorbs at approximately 350 nm and the ketoenamine form at 420 nm. After buffer exchange of the enzyme to remove any free PLP, the TamD PLP absorption peaks can be seen at 341 and 428 nm (Figure 3.20a). In its resting state, the TamD PLP equilibrium lies towards the ketoenamine form though a peak can still be observed for the enolimine isoform. Varied concentrations of L-serine (0-80 mM) were subsequently added to the enzyme and the change in the UV/Visible spectrum monitored.

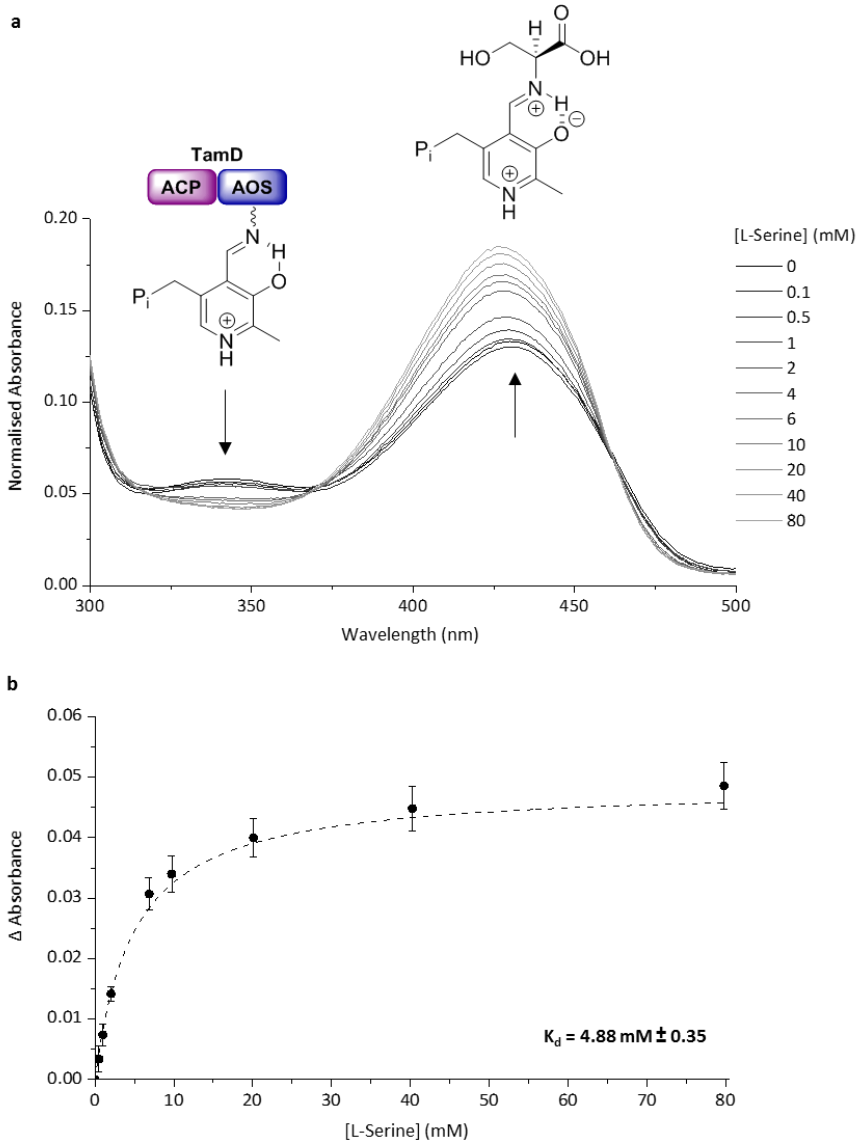


Figure 3.20: (**a**) UV/visible spectrum of TamD with varying concentrations of L-serine (0-80 mM) showing an increase and shift in the 428 nm peak to 426 nm and a decrease in the 341 nm peak with increasing concentrations of L-serine and (**b**) plot of the change in absorbance at 426 nm to calculate the dissociation constant (K_d) for L-serine (4.88 mM ± 0.35) with error bars indicating standard deviation of three repeats.

When L-serine is added to TamD, its PLP absorption spectrum shifts further towards the ketoenamine and away from the enolimine, so a concurrent decrease in the 341 nm peak and an increase and slight shift in the 428 nm peak to 426 nm is observed. The change in the UV/visible spectrum at different L-serine concentrations can then be used to calculate the dissociation constant (K_d) of TamD for L-serine (Figure 3.20b). This is calculated by plotting the change in the absorbance at 426 nm for each L-serine concentration. The points are fitted with a hyperbolic saturation curve and the asymptote of this curve is ΔA_{max}. The K_d is calculated as the concentration at ½ ΔA_{max} according to the equation in Section 6.8.1. The K_d for L-serine was determined as 4.88 mM ± 0.35, which is in a similar range to the other AOS enzyme which utilizes L-serine, SPT, whose homologues typically have a K_d of approximately 1 mM.[93] Binding of other amino acids to TamD was also tested, starting with the substrate enantiomer, D-serine.

Interestingly, D-serine has a slightly lower calculated dissociation constant than L-serine for TamD at just 3.97 mM ± 0.30. Though this is unexpected for an AOS to bind the D-enantiomer with similar affinity to the L-enantiomer, some AOSs, like the *M. tuberculosis* AONS can bind and turn over both L- and D-alanine.[185] Therefore, it is possible that TamD can use both enantiomers to produce HBM.

In addition to D-serine, five other amino acids were combined with TamD to determine their dissociation constants (L-alanine, glycine, L-threonine, L-aspartic acid and L-asparagine). If TamD is promiscuous and able to bind a variety of substrates, other unnatural amino acids may also be accepted in the reaction. This would result in the formation of novel bipyrroles which could be interesting building blocks for useful organic molecules. The tested amino acids produced no significant changes in the PLP spectra suggesting poor binding, so much so that the dissociation constants could not be calculated. For L-asparagine, the higher concentrations of the amino acid caused the protein to precipitate, resulting in the changes observed across the spectrum. The amino acid binding reactions indicate that TamD may have a narrow substrate specificity. This could be necessary in the cellular environment to ensure that only HBM is produced, as opposed to other bipyrroles which cannot be further modified by downstream Tam enzymes.

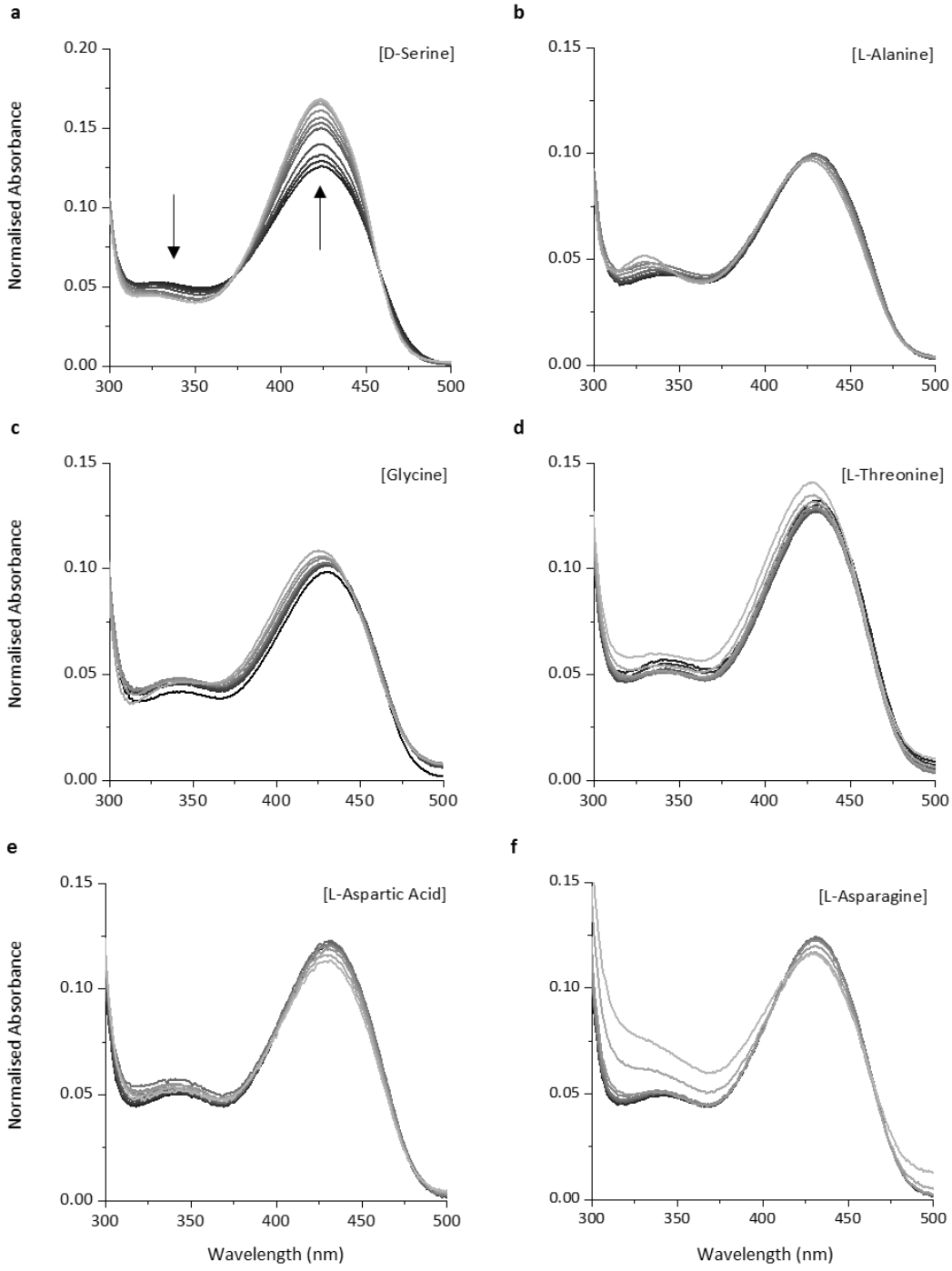


Figure 3.21: UV/visible spectra of TamD with varying concentrations of (**a**) D-serine, (**b**) L-alanine, (**c**) glycine, (**d**) L-threonine, (**e**) L-aspartic acid and (**f**) L-asparagine.

3.6 *In vitro* HBM Biosynthesis

MS analysis of each of the proteins and covalently bound intermediates involved in HBM biosynthesis has shown that the combination of TamF, pyrrole SNAC, TamD ACP, *E. coli* FabD and malonyl-CoA produce diketopyrrole-TamD ACP as in Figure 3.22. UV/visible analysis of the TamD protein in combination with its predicted amino acid substrate L-serine, indicated that this amino acid binds with the expected affinity to the AOS PLP domain. Therefore, the final step in the biosynthesis of HBM can be carried out, which is the second Claisen-like condensation of L-serine with the ACP-bound diketopyrrole thioester (Figure 3.22). This should release a linear product which subsequently cyclizes to yield HBM.

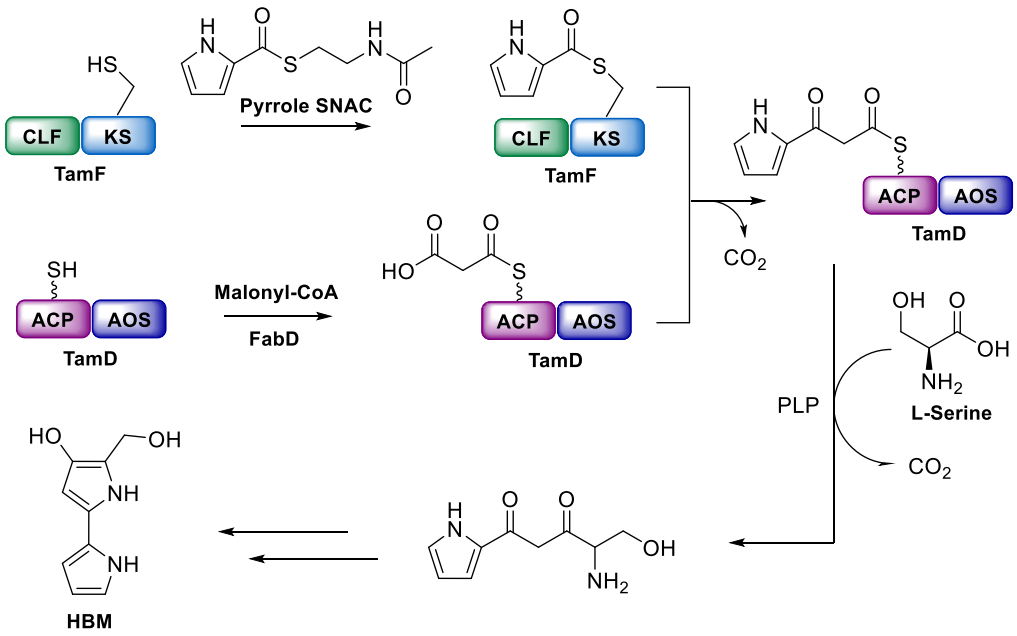


Figure 3.22: *In vitro* biosynthesis of HBM using TamF, TamD, *E. coli* FabD, pyrrole SNAC, malonyl-CoA and L-serine.

In the analogous Pig system from *Serratia marcescens,* Garneau-Tsodikova *et al*. were also able to follow the initial intermediates of these reactions *via* mass spectrometry or high performance liquid chromatography (HPLC) techniques, but they were unable to detect the final product, HBM.[55] As such, HBM has never previously been detected in an *in vitro* reaction. One drawback to the reaction method used by Garneau-Tsodikova *et al*., was that malonyl groups were installed directly onto the ACP with Sfp, so the enzymes could only undergo one turnover, limiting the HBM concentration. Thus a different approach was developed overcoming this limitation with the use of the *E. coli* MAT FabD.

Another potential limitation is that HBM is a bipyrrole and is likely to have a higher absorbance than those typically used for HPLC. To pre-empt this problem, the reaction was initially monitored by scanning the UV/visible spectrum from 320 nm to 480 nm every 30 sec (Figure 3.23a). The scans showed a peak clearly increasing over time with a λ_{max} of approximately 345 nm. The exact λ_{max} is not clear as other components of the reaction begin to interfere in the spectrum below this wavelength.

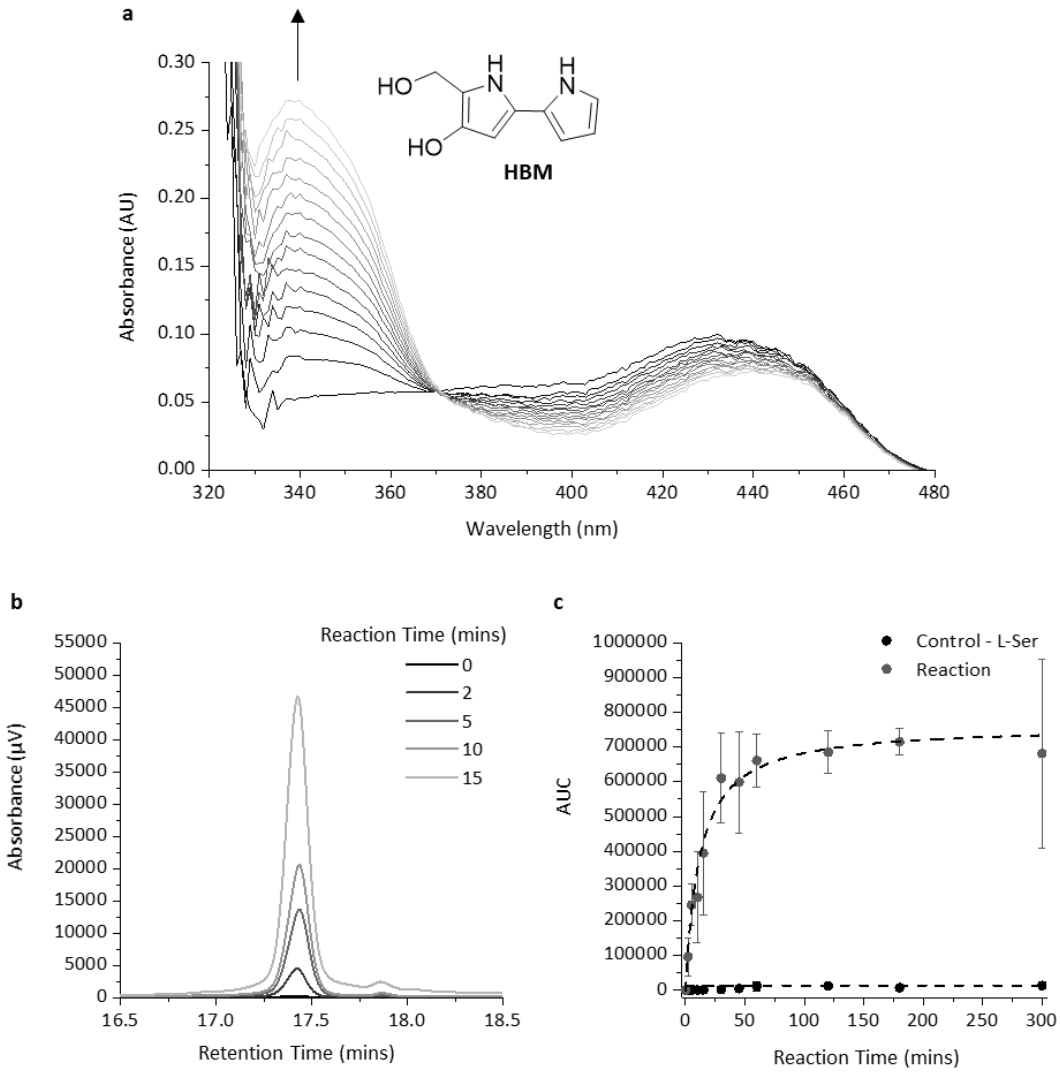


Figure 3.23: (**a**) UV/visible spectra of the HBM producing reaction containing TamF, TamD, *E. coli* FabD, pyrrole SNAC, malonyl-CoA and L-serine, at 30 sec intervals showing an increase in a peak at 345 nm, (**b**) HPLC traces at 345 nm showing the HBM peak increasing over time and (**c**) area under curve (AUC) of the HBM peak versus time for the control (without L-serine) and full reaction showing that the reaction has almost reached a plateau at 60 min. The error bars indicate the standard deviation of three repeats.

As a result of the UV/visible spectrum of the reaction, HPLC of the quenched reaction mixture was carried out using a reverse phase C18 column monitoring the eluent at a wavelength of 345 nm at different time points (0-5 h). Over this period, a new peak was observed at 17.4 min which appeared to correspond to the HBM product (Figure 3.23b). The reactions were repeated in triplicate and the area under the curve (AUC) calculated for each time point (Figure 3.23c) which produced a clear hyperbolic curve that reached a plateau around 60 min.

Since an analytical standard was not available for the HBM molecule, the HPLC peak believed to correspond to HBM, was analyzed by LC ESI-MS using a MicroTOF MS (Figure 3.24). An *m/z* of 179.0794 was detected, which is consistent with the predicted *m/z* of 179.0821 for $C_9H_{11}N_2O_2$ (error 15 ppm). This provides very strong evidence that the peak eluting at 17.4 min with an absorbance maximum of 345 nm is indeed HBM.

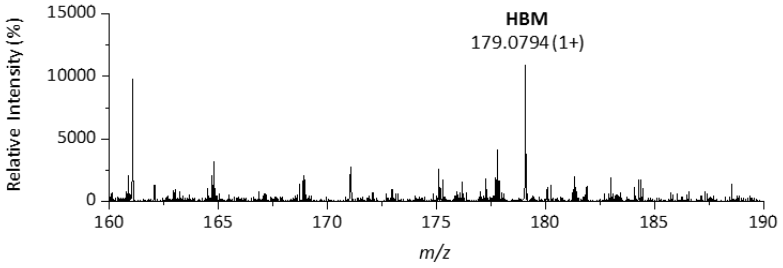


Figure 3.24: ESI-MS spectrum of the peak produced in the HBM biosynthetic reaction absorbing at 345 nm containing the exact *m/z* of HBM (179.0794) which is consistent with the predicted *m/z* of 179.0821 for $C_9H_{11}N_2O_2$ (error 15 ppm).

3.7 Amino Acid Specificity of HBM Biosynthesis

Once HBM production could be routinely monitored by HPLC, the amino acid specificity of TamD was probed to see if it matched that of the binding assays or whether it has wider substrate promiscuity. Reactions were set up in the same way as previously, substituting L-serine with other amino acids at the same concentration. If successful, these reactions would produce the alternative bipyrrole products displayed in Figure 3.25. L- and D-serine would both produce HBM as the stereochemistry at the α-carbon is lost when the linear product is deprotonated upon cyclisation.

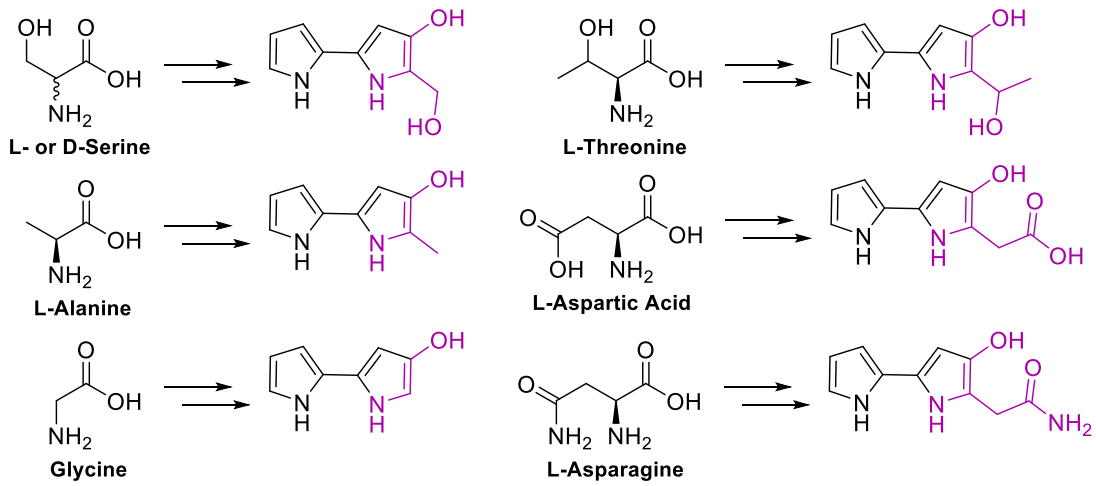


Figure 3.25: The bipyrrole HBM analogues that would be produced with alternative amino acid substrates in the bipyrrole producing reaction (Figure 3.22). Amino acid incorporation into the product is indicated in purple.

A peak was observed at the same retention time for the natural substrate (L-serine) when using D-serine (Figure 3.26a) though the new peak had a smaller area than the results from reaction with L-serine after 5 h reaction time (Figure 3.26b). The MS spectra also show that a product with the same mass is produced for both L- and D-serine (Figure 3.27a,b).

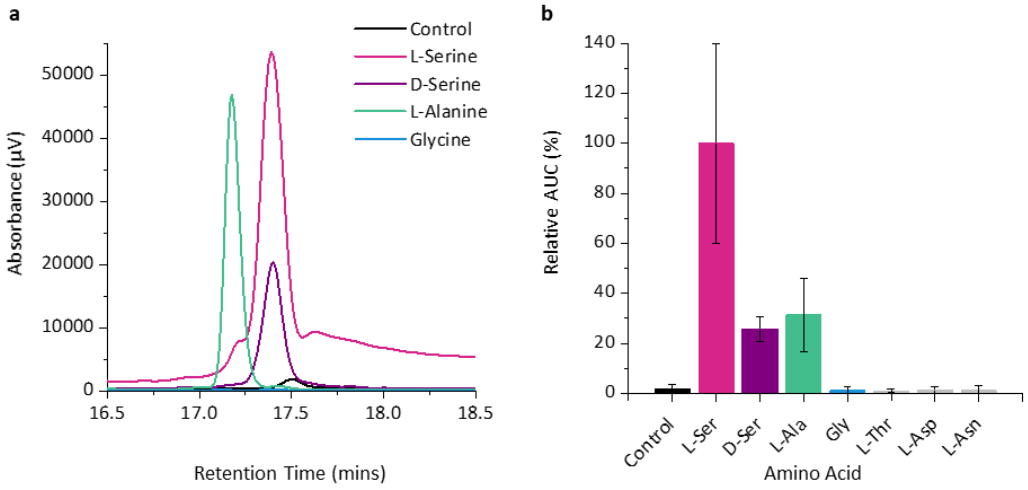


Figure 3.26: (**a**) HPLC traces at 345 nm of the reactions between TamF, TamD, *E. coli* FabD, pyrrole SNAC, malonyl-CoA and alternative amino acid substrates after 5 h incubation and (**b**) relative percentages of the area under the curve (AUC) compared to the native substrate (L-serine) and a control without amino acid. Error bars indicate standard deviation of three repeats.

However, the ability of this enzyme to accept both enantiomers is unusual with regards to its mechanism. All AOS enzymes studied thus far, first deprotonate the amino acid substrate at the α-carbon before catalyzing attack of the thioester (Section 1.5.2).[142,157,186] This deprotonation only occurs when the proton on the α-carbon is perpendicular to the plane of the PLP-ring forming a Dunathan intermediate.[136] Therefore the D-enantiomer often acts as an inhibitor for the enzyme as it cannot bind in the required orientation. However, *M. tuberculosis* AONS is also able to turn over L- and D-alanine so some AOS enzymes accept both enantiomers.[185] A structure of TamD would be helpful to rationalize this observation as it would indicate how these enantiomers bind and are deprotonated in the active site.

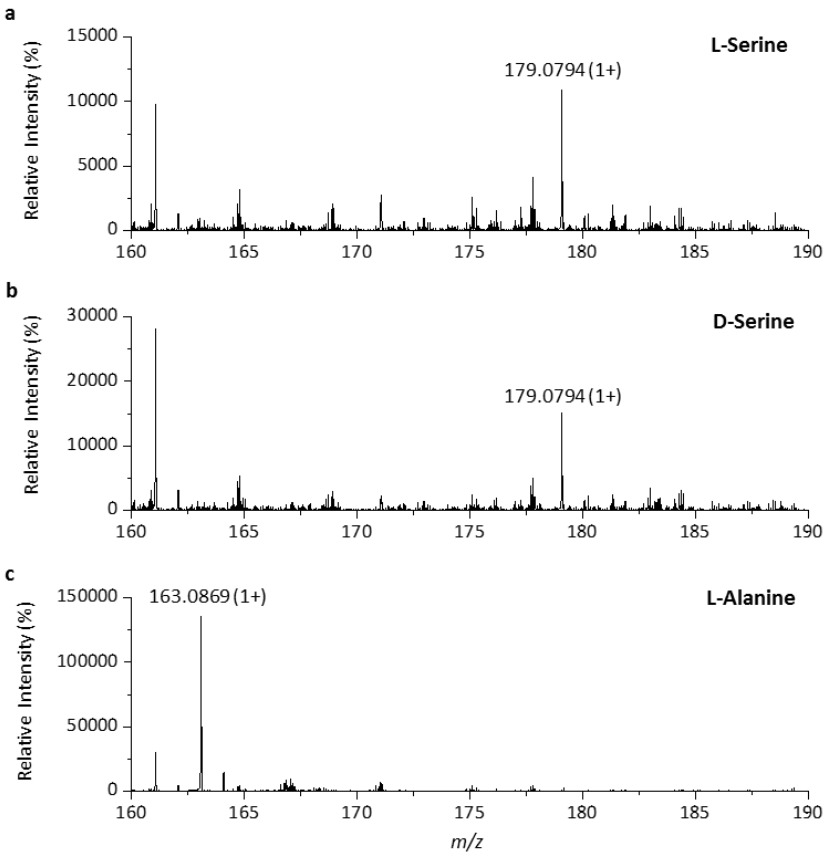


Figure 3.27: ESI-MS spectra of HBM analogues produced with (**a**) L-serine (**b**) D-serine and (**c**) L-alanine.

Despite being unable to calculate a K_d for L-alanine in the binding assays, this amino acid appears to be utilized by TamD. A new peak is observed in the HPLC trace at 17.2 min (Figure 3.26a), eluting just ahead of HBM. The average relative peak area is ~30% of the L-serine peak which suggests that L-alanine is reasonably well accepted in this reaction. MS analysis of the peak does not show the HBM *m/*z but instead the predicted *m/z* for the bipyrrole

without the hydroxyl group is detected (predicted mass for $C_9H_{11}N_2O$: 163.0871 Da, observed: 163.0869 Da, Figure 3.27c). Glycine, L-threonine, L-aspartic acid and L-asparagine are not substrates for TamD as no other peaks were found in the HPLC traces. This agrees with the observations in the binding assays, suggesting that TamD does have a very narrow substrate specificity, probably in order to biosynthesize tambjamine YP1 with high fidelity.

3.8 Pyrrole Specificity of HBM Biosynthesis

In order to produce tambjamine analogues which may have improved or more specific biological activity, the TamF/TamD reaction system will need to accept different heterocyclic substrates. To probe the specificity of the TamF and TamD enzymes for the first heterocycle, synthesis of alternative SNAC molecules was undertaken. Several alterations to the heterocycle moiety of the SNACs were made: substitution of the pyrrole ring to a furan, expanding the ring size by one carbon to a pyridine, adding a halogen substituent to the pyrrole and finally substituting the pyrrole for an expanded indole ring system (Figure 3.28). The indole analogue was chosen because obatoclax, an indole derivative of this molecule is an important anti-cancer drug.[187]

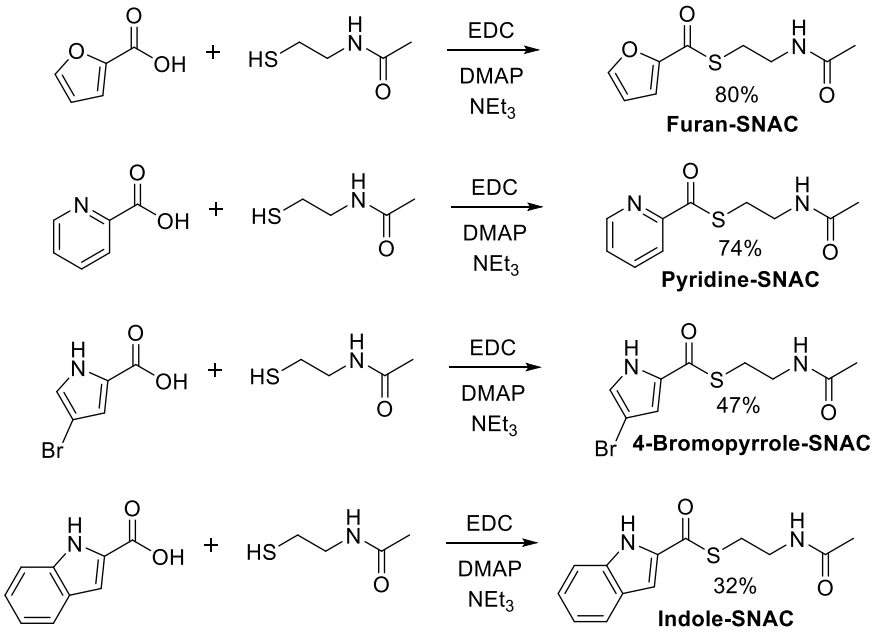


Figure 3.28: Synthesis of alternate SNAC molecules from carboxylic acids and *N*-acetylcysteamine to test the specificity of the TamF/TamD reaction system.

Three of the carboxylic acid substrates for the coupling reaction were commercially available, only the 4-bromopyrrole derivative was only available as an ethyl ester. The ester was hydrolyzed using potassium hydroxide in methanol and recovered in 94% yield. Coupling of the SNACs was performed using the same method as previously described for pyrrole-SNAC (NMR and MS data can be found in Appendix 12-15). Despite the 4-bromopyrrole and indole SNACs containing small impurities, the desired molecules were synthesized and it was assumed that these impurities would not affect the enzymatic reaction.

The four new SNACs were tested for binding to TamF by LC ESI-MS. When the natural pyrrole SNAC is combined with TamF, a shift from one single peak to another occurs as the pyrrole is loaded onto the active site cysteine (Figure 3.29a,b and deconvoluted masses in Table 3.1).

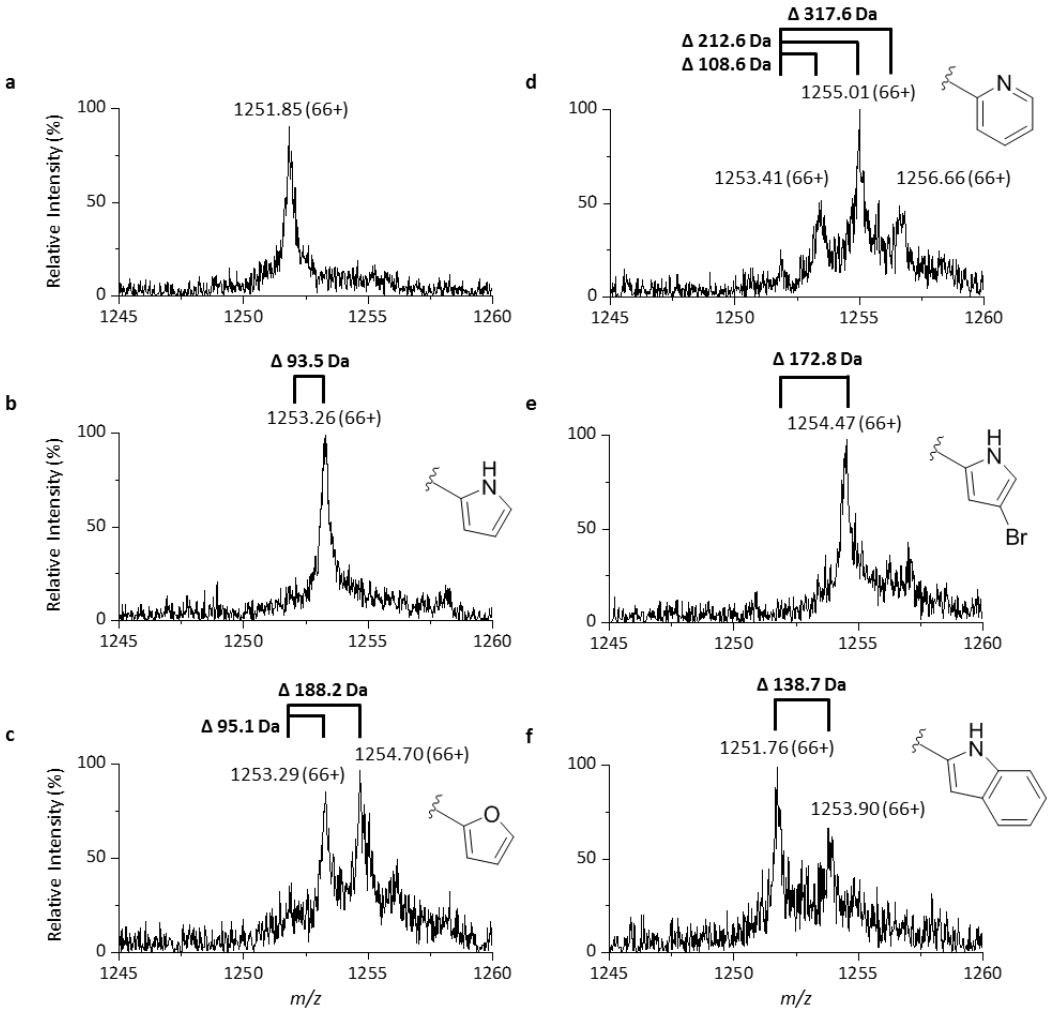


Figure 3.29: Denaturing LC ESI-MS spectrum of recombinant TamF showing the 66+ charge state (**a**) as purified, (**b**) with 1 mM pyrrole SNAC, (**c**) with 1 mM furan SNAC, (**d**) with 1 mM pyridine SNAC, (**e**) with 1 mM 4-bromopyrrole SNAC and (**f**) with 1 mM indole SNAC.

Protein	Predicted mass	Observed mass	Δ Predicted	Δ Observed
TamF	82554.3	82555.2+/-1.6		
TamF + pyrrole	82647.3	82648.7+/-1.2	93.1	93.5
TamF + furan	82648.3 82742.2	82650.3+/-1.9 82743.4+/-3.3	94.1 188.2	95.1 188.2
TamF + pyridine	82659.4 82764.5 82869.6	82663.8+/-2.3 82767.8+/-2.3 82872.8+/-2.8	105.1 210.2 315.3	108.6 212.6 317.6
TamF + 4-bromopyrrole	82726.3	82728.0+/-1.5	172.0	172.8
TamF + indole	82697.4	82693.9+/-1.9	143.1	138.7

Table 3.1: Predicted and deconvoluted observed masses and mass differences (Da) for modifications of TamF.

In contrast, when furan SNAC is employed as the substrate, two new peaks appear in the TamF spectrum (Figure 3.29c) corresponding to the addition of one and two furan moieties to the protein. This suggests that furan SNAC is reactive enough to modify other cysteine or lysine residues on the protein surface. Without further analysis of the protein, the positions of these moieties are unclear and may not include the active site cysteine. The same pattern is also observed for pyridine SNAC (Figure 3.29d). This substrate is even more reactive as higher intensity is observed for the twice modified protein than that of the once modified. The 4-bromopyrrole substrate modifies TamF only once, with a very small peak which may be a twice modified protein (Figure 3.29e). For the indole substrate, two peaks are detected, the first is the unmodified protein which is higher in intensity than the modified protein peak (Figure 3.29f).

The ability of the alternate SNACs to modify the TamF protein follows the order of increasing reactivity: indole < pyrrole < 4-bromopyrrole < furan < pyridine. This trend is aligned with the chemical reactivity of these molecules and may not be dependendent on the enzymatic activity. As these molecules have aromatic rings conjugated to the thioester carbonyl, their reactivity increases with decreasing resonance stability. This is likely to be the reason why the furan and pyridine SNACs are able to modify the protein multiple times.

In order to test whether the new heterocycles are binding at the TamF active site and can subsequently be turned over in a Claisen-like condensation, they were also incubated with

TamD ACP_{98}, malonyl-CoA and *E. coli* FabD and subjected to LC ESI-MS (Figure 3.30). The deconvoluted masses of each of the TamD ACP modifications can be found in Table 3.2.

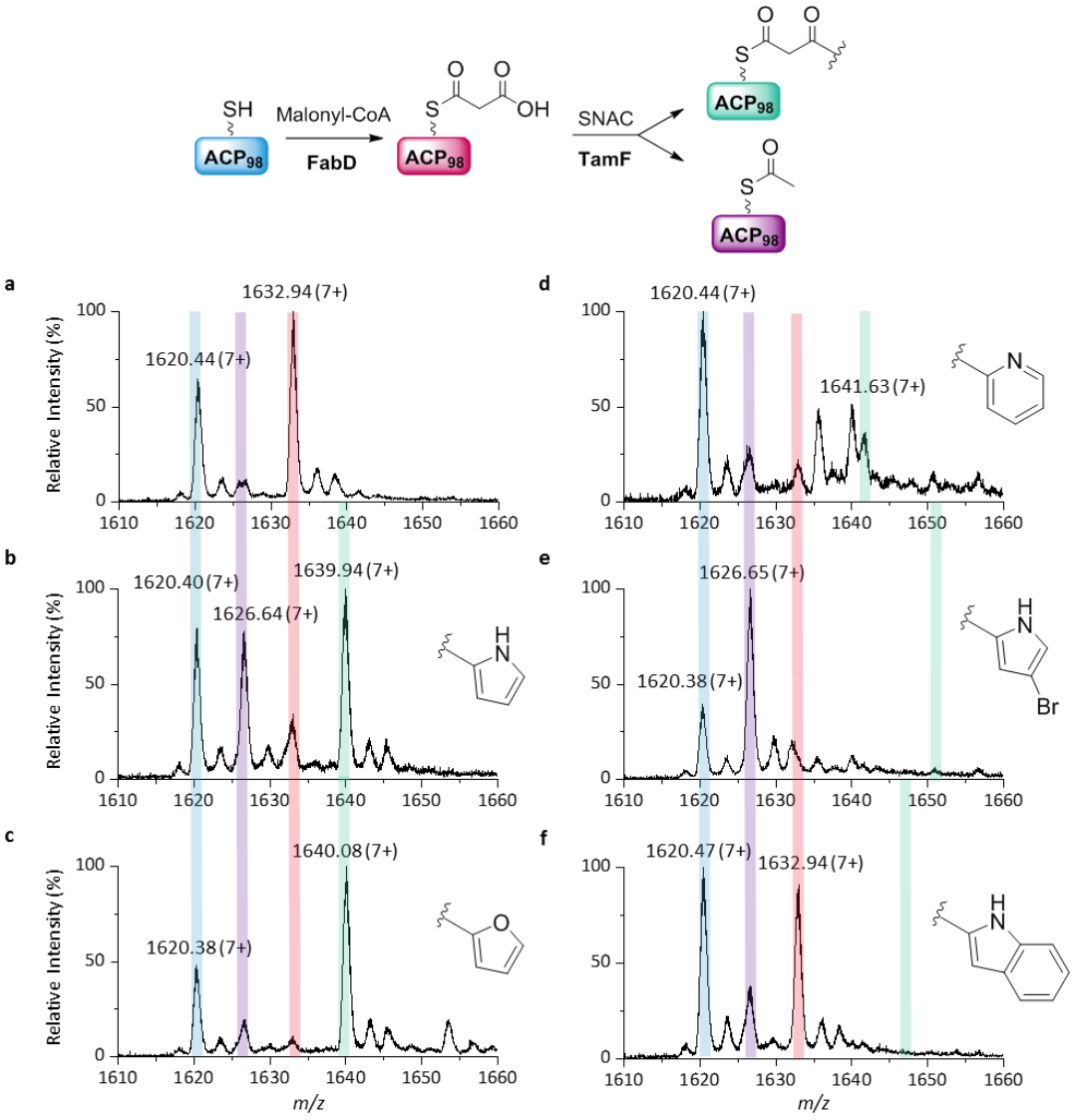


Figure 3.30: Denaturing LC ESI-MS spectrum of recombinant *holo*-TamD ACP_{98} showing the 7+ charge state after 15 min incubation with malonyl-CoA, *E. coli* FabD and TamF (**a**) only, (**b**) with pyrrole SNAC, (**c**) with furan SNAC, (**d**) with pyridine SNAC, (**e**) with 4-bromopyrrole SNAC and (**f**) with indole SNAC.

In the control reaction (Figure 3.30a), *holo*-TamD ACP is incubated with malonyl-CoA, FabD and TamF to show that without binding of a SNAC substrate to the TamF active site cysteine, decarboxylation does not occur, as no acetyl-ACP is observed. Though the furan and pyridine based substrates are able to react with malonyl-TamD ACP_{98} in the Claisen-like condensation (product peaks in green Figure 3.30c,d), it appears that these SNACs are so reactive that other

new peaks corresponding to addition of extra heterocycles to the TamD ACP also appear (particularly for the pyridine based substrate).

Protein	Predicted mass	Observed mass	Δ Predicted	Δ Observed
holo-TamD ACP	11337.6	11336.6 ± 0.4		
holo-TamD ACP + malonyl	11423.6	11423.6 ± 0.1	86.0	87.0
holo-TamD ACP + acetyl	11379.6	11379.5 ± 0.1	42.0	42.9
holo-TamD ACP + pyrrole product	11472.7	11472.7 ± 0.1	135.1	136.1
holo-TamD ACP + furan product	11473.6	11473.5 ± 0.1	136.0	136.9
holo-TamD ACP + pyridine product	11484.7	11483.7 ± 0.5	147.1	147.1
holo-TamD ACP + 4-bromopyrrole product	11551.6	Not observed	214.0	NA
holo-TamD ACP + indole product	11522.7	Not observed	185.1	NA

Table 3.2: Predicted and deconvoluted observed masses and mass differences (Da) for modifications of TamD ACP_{98}.

In contrast, the 4-bromo and indole substrates are not transferred to the ACP by the Claisen-like condensation. However, they do induce decarboxylation of malonyl-TamD ACP illustrated by the disappearance of the malonyl peak (pink) and appearance of the acetyl peak (purple). For the bromo-substituted pyrrole, the decarboxylation is complete and no more malonyl-ACP can be observed, whereas, the indole substrate only induces partial decarboxylation consistent with fully and partially modified TamF respectively. These data illustrate that both the bromo and indole substrates are bound in the active site of TamF (as they are able to induce decarboxylation[105]) but are not captured by the TamD ACP_{98} carbanion after decarboxylation. The lack of 4-bromopyrrole condensation product is not in line with the reactivity of this thioester which should be slightly more reactive than the pyrrole thioester. Therefore, the absence of product may be due to the steric hindrance of the bromine substituent which inhibits the attack of the TamD ACP carbanion. Indeed, though some bromine substituted tambjamines have been identified, none of these contain the bromine in the 4-position. For the indole substrate, along with its inherent stability, steric

hindrance is likely to be a contributing factor for its low reactivity. This may prevent it from entering the TamF active site as well as inhibit attack of the TamD ACP carbanion. The TamF/TamD system would have to be engineered to accept bulkier substrates like the 4-bromopyrrole and indole derivatives.

The furan and pyridine SNAC substrates were also combined in reactions with the other components required for HBM production (TamF, TamD, *E. coli* FabD, malonyl-CoA and L-serine). The reactions were monitored both by UV/visible spectroscopy and HPLC, however, neither of these produced the expected novel bicyclic systems. It is surprising that these substrates were not turned over by TamD, as the first Claisen-like condensation product could be observed. Since these molecules are so reactive and are able to modify the proteins unspecifically, it may be that the surface modifications inactivate TamD and thus no product can be formed. In this case, utilizing the preceding enzymes in the pathway and feeding the carboxylic acid substrates may be necessary to observe product formation.

3.9 HBM Biosynthesis with Unfused TamD Domains

As the TamD ACP domain was successfully truncated from the AOS domain, the AOS was also truncated from the ACP to reconstitute the activity of the two unfused domains. Fusion of an ACP and an AOS is unusual so this may be necessary due to the low binding affinity of the two domains for one another. A construct produced by Daynea Wallock-Richards without the first 95 residues of TamD (the ACP domain) left the linker region and AOS domain in pET22b with a C-terminal 6xHis tag (TamD $AOS_{\Delta 95}$, sequence in Appendix 16). This protein was expressed in the same way as TamD in 2 L of *E. coli* BL21 (DE3) cells with 0.1 mM IPTG at 20 °C for 4 h. The protein was purified by nickel IMAC and GFC (Figure 3.31), resulting in a broad peak eluting at 63.9 mL consistent with the expected protein dimer (calibration curve in Appendix 5). The final yield of this protein was excellent at 27 mg/L culture.

LC ESI-MS showed TamD $AOS_{\Delta 95}$ to have a deconvoluted mass of 49278.7 Da ±1.3 Da (Figure 3.32) consistent with the calculated mass of 49268.7 Da. A smaller secondary peak can also be seen at each of the charge states with a mass increase of 230.3 Da, the calculated mass of PLP, which is still attached to some of the denatured protein.

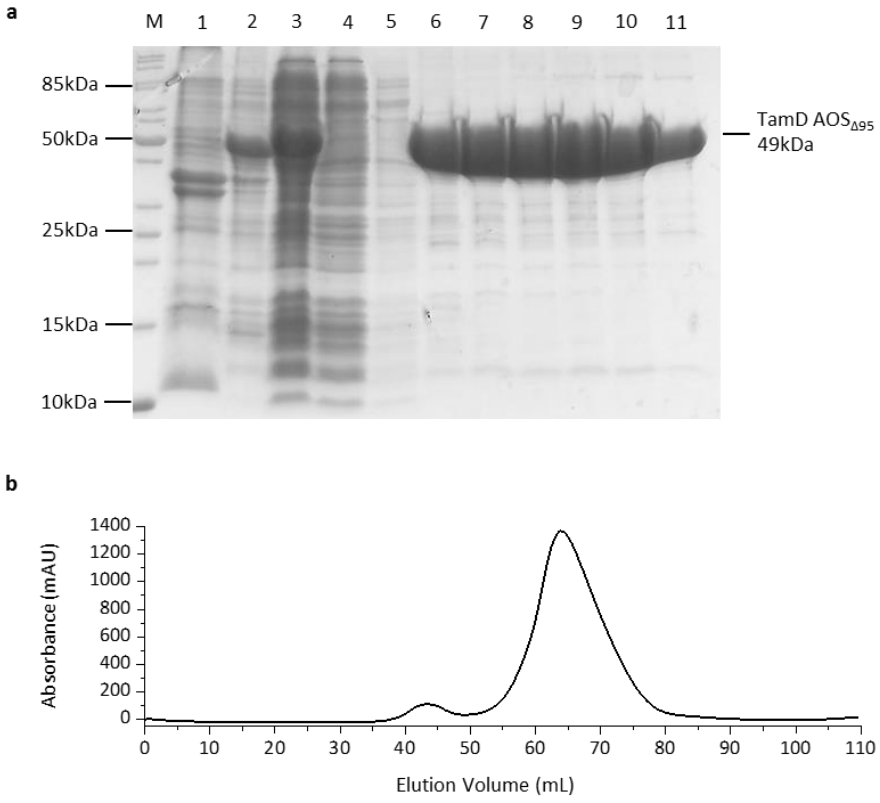


Figure 3.31: TamD AOS$_{\Delta 95}$ purification (**a**) SDS-PAGE analysis: (M) marker, (1) insoluble and (2-3) soluble fractions of cell pellet, (4-11) 3 mL fractions from the Superdex S200 GFC, (**b**) chromatogram of Superdex S200 GFC showing peak elution at 63.9 mL, indicative of a protein dimer.

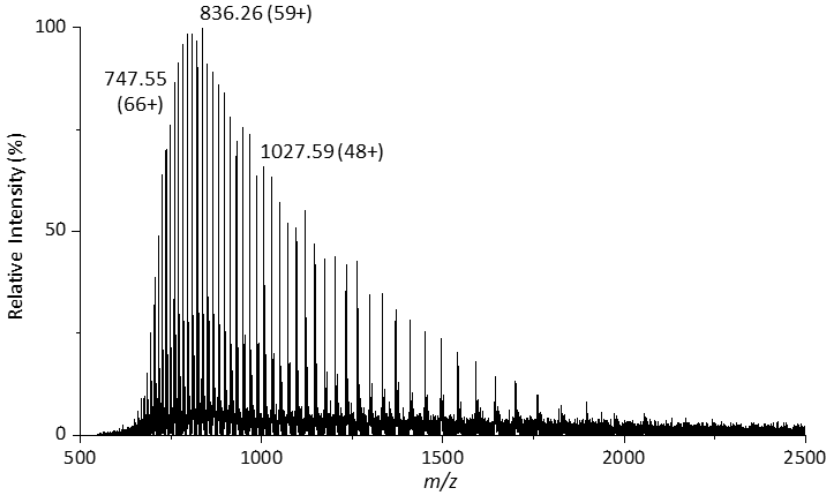


Figure 3.32: Denaturing ESI-MS spectrum of recombinant TamD AOS$_{\Delta 95}$ with a deconvoluted mass of 49278.7 Da ± 1.3 Da.

The K_d of the TamD $AOS_{\Delta 95}$ for L-serine was measured to ensure it had similar properties to the full-length enzyme. The K_d value for the AOS domain was the same order of magnitude as the full-length protein at 3.1 mM ± 0.1 (Figure 3.33).

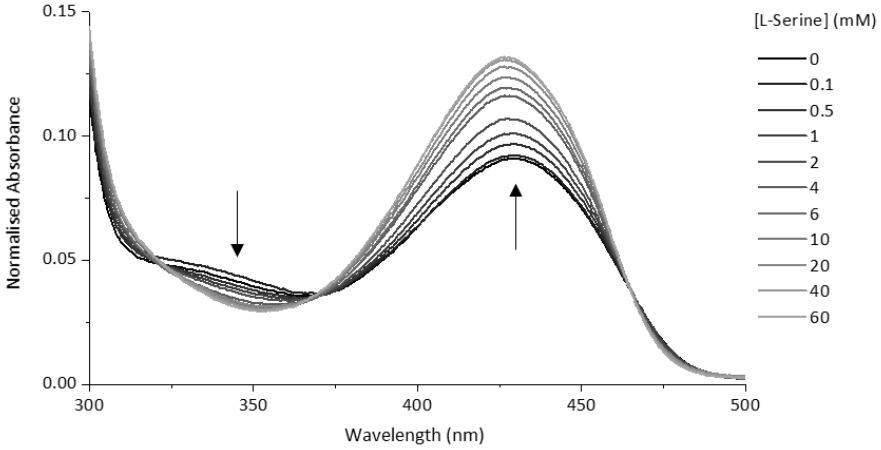


Figure 3.33: UV/visible spectrum of TamD $AOS_{\Delta 95}$ with varying concentrations of L-serine (0-60 mM) showing an increase in the 426 nm peak and a decrease in the 347 nm peak with increasing concentrations of L-serine.

The TamD $AOS_{\Delta 95}$ subunit was subsequently used in the HBM assay with TamD ACP_{98} to determine whether the separated subunits are able to work together to produce HBM. The reaction was also monitored by HPLC (Figure 3.34a) showing that it does produce a peak at the same retention time (17.4 min) as the reaction with the full-length protein. However, analysis of the AUC (Figure 3.34b) shows that much less HBM is produced over the 5 h time period when using the two independent domains. Fusion of the domains clearly increases the rate of reaction as it holds the two domains in place and reduces the freedom of movement. Gene fusion is not only used by nature to improve product formation but has also been used biocatalytically to improve turn over and rates of reaction.[188]

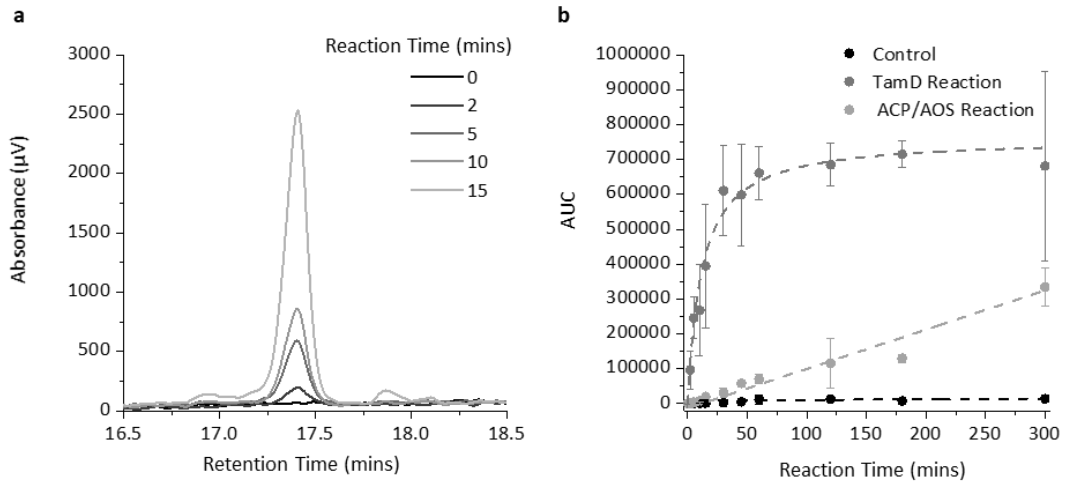


Figure 3.34: Reactions for producing HBM with TamD ACP_{98} and TamD $AOS_{\Delta 95}$ in place of full-length TamD containing TamF, *E. coli* FabD, pyrrole SNAC, malonyl-CoA and L-serine (**a**) HPLC traces at 345nm showing the HBM peak increasing over time and (**b**) area under curve (AUC) for the HBM peak versus time for the reaction with the full-length protein as well as the separated subunits and the control without L-serine. The error bars indicate standard deviation of three repeats.

3.10 HBM Protein Structural Studies

3.10.1 TamF X-ray Crystallography

TamF is one of only a few characterized KS-CLF dimers and is particularly interesting as it is a fusion protein which may form an $\alpha_2\beta_2$ tetramer. Therefore, it would be very interesting to obtain a crystal structure of this enzyme to shed light on the CLF mechanism of action. TamF was prepared for crystal trials by purification as previously reported, followed by concentration of the protein to 12 mg/mL by Vivaspin membrane. Crystallization of this protein mixture was subsequently attempted in five 96-well plate sitting drop screens prepared using a Gryphon robot (in collaboration with Dr. Stephen McMahon and Prof. Jim Naismith, University of St. Andrews). Two of the screens were the commercially available WIZARD and JCSG+, along with three bespoke University of St. Andrews polyethylene glycol (PEG) based screens PEGS1-3. From the five screens, two hits were identified for TamF crystals: PEGS1 A9 - 51.12% PEG400, 100 mM HEPES pH 7.5, 40 mM magnesium formate and PEGS2 H9 - 34% PEG400, 100 mM MOPS pH 6.5, 120 mM calcium acetate, 0.02% BOG. Though these crystals were small and amorphous, optimization plates were nonetheless

created from these conditions, varying the pH and concentration of each of the additives. However, all the crystals produced were of a similar shape and size and the best diffracting crystal was only able to diffract to 20 Å.

The crystals from PEGS1 H9 were subsequently used as seed for four new screens (PEGS1-3 and JCSG+). These screens produced more hits with larger and more defined crystals; the best conditions identified are summarized in Table 3.3. Five crystals from each of these wells were shot on the St. Andrews X-ray source however, the best diffracting crystal diffracted to only 7.5 Å. As this is not enough to characterize the structure of a protein, no further crystallography was attempted for TamF.

Screen	Well	Conditions
JCSG+	A12	20% (w/v) PEG3350, 200 mM KNO_3
PEGS1	B12	34% (w/v) PEG3350, 100 mM bicine pH 9.5, 30 mM ammonium tartrate
PEGS3	C2	24% (w/v) PEG4000, 100 mM CHES pH 9.0, 300 mM LiCl

Table 3.3: Conditions of TamF crystals that were shot on the University of St. Andrews X-ray source.

3.10.2 Full-length TamD X-ray Crystallography

TamD is also an interesting target for crystallography as it is an unusual ACP-AOS didomain and it is one of only a few AOS domains that accept aromatic substrates. A structure of this enzyme would not only give an insight into the interaction between an ACP and an AOS domain but it may also shed light on the specificity of AOS enzymes in general. Therefore, TamD crystal trials were carried out by the same method as TamF. The protein was concentrated to 10 mg/mL and sitting drop crystal trials were set up using a Gryphon robot. Eight different screens were utilized: WIZARD, JCSG+, PEGS1-4 and two other bespoke University of St. Andrew's screens, SA19 and SA20. For TamD there were several conditions which produced crystals however, these were all thin needle-like crystals which are difficult to mount for crystallography. The best conditions for the TamD crystals are outlined in Table 3.4. Despite optimizing around all of these conditions and using these crystals for seeding in new screens, no improved crystals were obtained.

Screen	Well	Conditions
WIZARD	F6	20% (w/v) PEG3000, 100 mM Tris/HCl pH 7.0, 200 mM calcium acetate
JCSG+	A5	20% (w/v) PEG3350, 200 mM magnesium formate
PEGS1	A6	20.29% (w/v) PEG4000, 100 mM MOPS pH 7.0, 110 mM magnesium formate
PEGS1	A8	12.6% (w/v) PEG8000, 100 mM HEPES pH 8.0, 190 mM KCl
PEGS3	G6	33.19% (w/v) PEG3350, 100 mM MOPS pH 7.0, 120 mM $CaCl_2$

Table 3.4: Conditions of the best TamD crystals that were used to further optimize the crystallization conditions.

3.10.3 TamD AOS X-ray Crystallography

The presence of a flexible linker region between the two TamD domains could explain the difficulty observed in obtaining TamD crystals. Since the two domains are dynamic with respect to one another, this could make uniform crystal packing less accessible. Therefore, TamD $AOS_{\Delta 95}$ crystal trials were also carried out without its ACP partner.

As before, eight screens were set up: WIZARD, JCSG+, PEGS1-4 and SA19-20. Though more conditions were identified containing $AOS_{\Delta 95}$ crystals than for the full-length protein, these were still needles. However, some needles were longer and thicker so the best conditions (Table 3.5) were shot on the St. Andrews X-ray source.

Screen	Well	Conditions
WIZARD	F6	20% (w/v) PEG3000, 100 mM Tris HCl pH 7.0, 200 mM calcium acetate
PEGS1	A8	12.6% (w/v) PEG8000, 100 mM HEPES pH 8.0, 190 mM KCl
PEGS3	G6	33.19% (w/v) PEG3350, 100 mM MOPS pH 7.0, 120 mM $CaCl_2$
PEGS3	H5	18.29% (w/v) PEG3350, 100 mM MOPS pH 6.5, 220 mM Li_2SO_4
SA20	A5	22.78% (w/v) PEGMME2k, 100 mM MES pH 6.0, 130 mM Li_2SO_4
SA20	G2	14.69% (w/v) PEGMME5k, 100 mM sodium cacodylate pH 7.0, 70 mM calcium acetate

Table 3.5: Conditions of TamD $AOS_{\Delta 95}$ crystals produced in Tris buffer that were shot on the University of St. Andrews X-ray source.

However, unfortunately, again none of these crystals produced any diffraction. The crystals were also colourless, suggesting a lack of PLP which results in bright yellow protein. TamD and its subunits were purified in Tris buffer, an amine that binds to PLP and can remove the cofactor from proteins that do not bind it tightly. Therefore, the buffer was swapped from Tris to 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) to combat this potential problem and determine if it has an effect on crystal formation.

Six crystal trials were set up with the protein in HEPES buffer: WIZARD, JCSG+, PEGS1-4. Despite changing the buffer, the crystals appeared very similar in the new screens. Nonetheless, crystals from the conditions in Table 3.6 were also shot on the X-ray source but none of these diffracted either.

Screen	Well	Conditions
WIZARD	F6	20% (w/v) PEG3000, 100 mM Tris/HCl pH 7.0, 200 mM calcium acetate
JCSG+	D6	20% PEG8K, 50 mM Tris pH 8.5, 200 mM $MgCl_2$
PEGS1	D11	16% PEG3350, 100 mM MOPS pH 7.0
PEGS2	C10	28% PEG3350, 100 mM CHES pH 9.0, 180 mM potassium thiocyanate, 3% methanol
PEGS3	F11	27% PEG6K, 150 mM magnesium formate

Table 3.6: Conditions of TamD $AOS_{\Delta 95}$ crystals produced in HEPES buffer that were shot on the University of St. Andrews X-ray source.

Though the TamD $AOS_{\Delta 95}$ construct does not include the ACP domain which is flexible with respect to the AOS domain, it still contains the flexible and disordered linker region. The presence of the disordered region on its own may still affect the ability of the protein to crystallize. Therefore, new AOS constructs were created, attempting to cut back the linker region to where the AOS domain begins. As the AOS domain is predicted to begin at residue 131 according to BLAST analysis, four constructs were made of increasing length: $AOS_{\Delta 131}$, $AOS_{\Delta 128}$, $AOS_{\Delta 125}$ and $AOS_{\Delta 122}$ (sequences in Appendix 17).

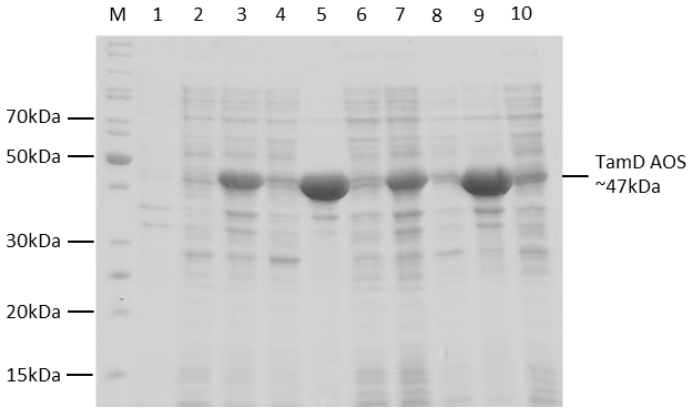


Figure 3.35: SDS-PAGE analysis of test expression of TamD AOS constructs: (M) Marker (1) insoluble and (2) soluble fractions of expression of empty plasmid, (3) insoluble and (4) soluble fractions of expression of TamD $AOS_{\Delta 122}$, (5) insoluble and (6) soluble fractions of expression of TamD $AOS_{\Delta 125}$, (7) insoluble and (8) soluble fractions of expression of TamD $AOS_{\Delta 128}$, (9) insoluble and (10) soluble fractions of expression of TamD $AOS_{\Delta 131}$.

Figure 3.35 shows that none of the constructs with the truncated linker region are soluble, as they all contain large bands in the pellet fraction and none in the supernatant. This gives some insight into the protein structure, suggesting that the linker region may have a more important function for the AOS domain than previously anticipated. However, due to the instability of these constructs, no further attempts at the crystallography of the TamD protein were attempted.

3.11 Conclusions

HBM has been successfully biosynthesized for the first time in an *in vitro* reaction using enzymes from the *tam* cluster. SNAC molecules were synthesized and used to load substrates onto the TamF protein without the need for the preceding enzymes in the pathway, allowing a more facile method of studying bipyrrole formation. The TamD protein and its ACP subunit were post-translationally modified with their critical 4'-PP arms by the PPTase Sfp, both *in vitro* and by *in vivo* coexpression. TamD was further loaded with its malonyl substrate using *E. coli* FabD in an innovative approach that resulted in the ability of this system to carry out more than one turnover. Finally, all the components of the HBM forming reaction were combined and an HPLC method was developed to monitor formation of the bipyrrole. This

has not only proven that HBM is formed in the tambjamine YP1 pathway but is also an important step in elucidation of the biosynthetic pathway of other natural products, such as the prodiginine family (Figure 1.2).

Both the amino acid and heterocyclic substrate specificity of the HBM forming reaction have also been explored. Though the TamD protein can accept L-alanine to form a novel bipyrrole, this protein appears to have very little promiscuity and would require engineering to accept other amino acids. Though no alternative heterocycles were accepted in the reaction system, utilizing these substrates has shed light on the constraints of the TamF catalyzed Claisen-like condensation. Indole and 4-bromopyrrole substrates appear to be too bulky to be turned over in the decarboxylative condensation. Whereas, the other alternative heterocycles (furan and pyridine) are too reactive in SNAC form so may inactivate the enzymes.

Though crystal structures of TamF and TamD were not obtained, some insight into the TamD fold has been gained through the cloning of various truncations. The expressed protein solubility indicates that the domain boundary for the ACP is around residue 98 of the TamD sequence and that the AOS domain requires its disordered linker region for stability. Further efforts to obtain a 3D structure of these enzymes will not only reveal the details of these interesting fusions but will also allow mutagenesis of these enzymes to accept a broader substrate range. This structure could be achieved by homology modelling or by crystallography, utilizing a homologous enzyme from an alternative organism.

4 Amine Biosynthesis

4.1 Introduction

The *P. tunicata* tambjamine YP1 biosynthetic pathway is predicted to be convergent, containing two fragments that are synthesized independently, then combined. The first is the bipyrrole moiety with an aldehyde (HBM, discussed in Chapter 3) which is ready to be condensed with a long chain amine. The biosynthetic route to the alkyl amine fragment was postulated by Burke *et al*.[50] to be initiated by the adenylation of a C12 fatty acid and its attachment to coenzyme A (CoA) as a thioester (Figure 4.1). The CoA thioester would be subjected to dehydrogenation by the flavin adenine dinucleotide (FAD)-dependent TamT enzyme installing the 3,4-alkene. Finally, the bifunctional TamH would reduce the CoA thioester and transaminate the product aldehyde to generate the final alkyl amine.

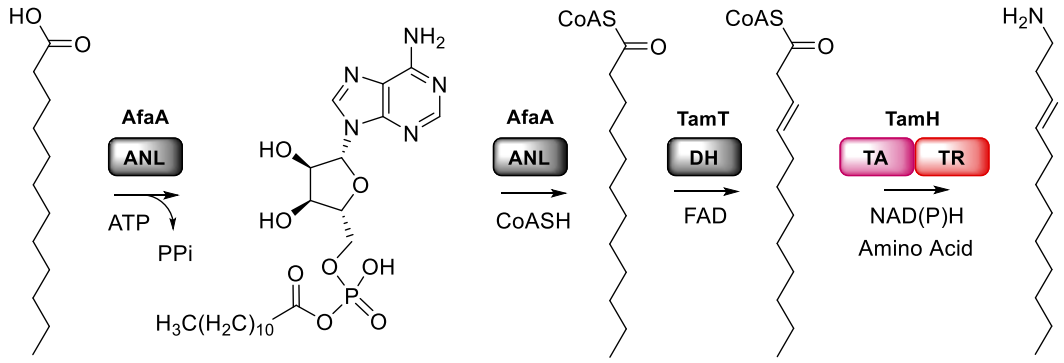


Figure 4.1: Predicted biosynthetic pathway to the alkyl amine of tambjamine YP1 adapted from Burke *et al*.[50] This pathway begins with the adenylation of C12 fatty acid by the ANL AfaA, followed by attachment of the fatty acid to CoA. The acyl thioester is likely to be dehydrogenated by TamT to give the 3,4-alkene and then reduced and transaminated by TamH to give the alkyl amine, though the order of the TamT and TamH reactions are unknown.

Burke *et al*. hypothesized that the activation reaction would be carried out by AfaA, an adenylation enzyme (ANL) that is a member of the fatty acid CoA ligase (FACL) family and is encoded outside the *tam* operon. FACL enzymes adenylate fatty acids using ATP to activate the carboxylic acid for nucleophilic attack by the terminal thiol of CoA.[189] Deletion of the *afaA* gene aborted biosynthesis of Tambjamine YP1. However, FACLs have many and varied roles in the cell, affecting fatty acid catabolism, protein acylation and even transcriptional regulation,[189] any of which could influence the biosynthesis of secondary metabolites.

The aim of this study is to elucidate the biosynthetic route to the tambjamine YP1 tail. Though AfaA could play a role in the biosynthesis, several open reading frames (ORFs) within the *tam* operon have not been designated functions within the biosynthetic pathway, so one of these could fulfill the role of activating C12 fatty acid for tambjamine YP1 biosynthesis. One in particular (*tamA*) is designated as an adenosine monophosphate (AMP)-ligating enzyme, so the enzyme encoded by this gene will be studied in more detail along with the other predicted enzymes that are involved in this pathway.

4.2 The TamA Protein

4.2.1 Bioinformatic Analysis

The gene encoding the predicted AMP-ligating enzyme TamA, was accessed from NCBI (2 kb, Appendix 18). BLAST analysis of the TamA protein sequence suggested that like several others in the operon, this 75 kDa enzyme contains two domains. The N-terminal domain showed highest similarity to FACL enzymes and the C-terminal domain to an acyl carrier protein (ACP, Figure 4.2, shown on the sequence in blue and green respectively).

```
MQCEASSLID LLAFDARNKP SQEVFRFVSD NGESEASYDY QTLSQEISRI AIGLQALIKT 60
SHNQDQALIV LPQGVQFVTA FYGCMAANVI AVPSFPPKSQ LQIERLQFAI TDLGNPIVIT 120
NRDILPKLQE HIALDSVRWL LIEDLASVIA QPLSDFRTHE HSIALLQYSS GTTGKPKGVI 180
ITNQNIMENS ELIRQSFGHK EDHTRMMLWL PPHHDMGLVG GVMQGVYTGY PTLLMPTDLF 240
LRSQYRWLKA VSDYRATTTG APNFAYELAV KNIRESRLAE LDLSSLENLF CGAEPINSHS 300
INQFLDKFAP CGLKPEAFLP CYGMAEATLM VSGKPHGQQY KQLCIDEPLL KHGMVKPLNT 360
PNAHSLWLVS SGVVHSSLQA RIVNPETGTE VAQGQVGEIW LQGSSISPGY WQDAERTAIN 420
FGLPLAGYEE TFHRTGDLGF YHQDELFITG RLKEVVIIRG ANFYPQDLEY ETTLAFPELN 480
NCRSAAFSVP KEGKEQLIMA IEVPRNVTEF NQYAKILNGR LVERFGIRAD IILFLPRKTI 540
KITSSGKLQR VAIKKAYEEQ QLPVYFQYQL QGEQIAPREV SLDISNQDSV AKWLVARVSE 600
LTGVAIAQIS EHEPLTNVGL DSVLAMEILF RLEQQTGVYL APDVLYSCNT PSLLAEQIIK 660
VAGNVAEKEL NLSC                                                  674
```

Figure 4.2: The TamA amino acid sequence (UniProtKB: A4C5W5) encoding a 75 kDa protein with the predicted adenylation domain (ANL, blue, 64 kDa) and the acyl carrier protein (ACP, green, 10 kDa) with the conserved DSX motif in bold.

As ACP is the protein analogue of CoA (detailed in Section 1.4.2), it is possible the TamA ANL domain could adenylate fatty acids for transfer to its attached ACP instead. Fatty acid AMP

ligases (FAALs) are a class of enzymes that transfer their fatty acids to ACP substrates rather than CoA.[190–193] They differ only slightly in structure to FACLs by the addition of an insertion loop that blocks CoA from the active site but is released upon ACP binding. In order to determine if the TamA adenylation domain falls into this category, it was aligned with known FAAL enzymes from the PDB (*Mycobacterium smegmatis* FadD32, 5D6J;[191] *Mycobacterium tuberculosis* FadD32, 5HM3;[192] *Mycobacterium marinum* FadD32, 5EY9;[193] *Legionella pneumophila* FAAL, 3KXW,[190] alignment in Appendix 19). It shares around 30% identity with these enzymes as well as containing the FAAL insertion loop. Therefore, TamA could be a FAAL which transfers C12 fatty acids to its fused C-terminal ACP domain (Figure 4.3), priming them for the downstream production of the amine tail. Alignment of the ACP with its closest homologues in the PDB (*Actinoplanes friulienses* LipD, 2N98;[194] *Burkholderia ambifaria* Bamb_5917, 5MTI and the *Aspergillus parasiticus* norsolorinic acid synthase (NSAS), 2KR5[195]) identified the conserved DSX motif where 4'-phosphopantetheine (4'-PP) modification occurs but showed little other similarity.

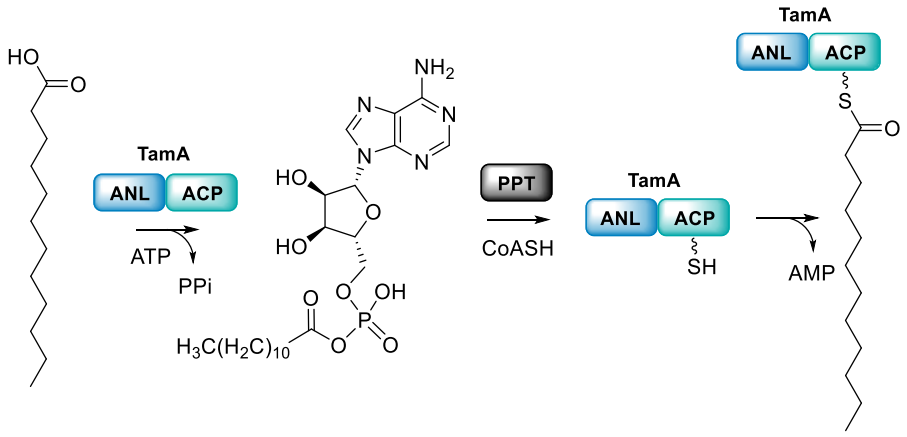


Figure 4.3: Predicted pathway of C12 fatty acid activation by TamA, beginning with adenylation of the fatty acid using ATP by the ANL domain, then post-translational modification of the TamA ACP and transfer of the fatty acid to the phosphopantetheine (4'-PP) thiol.

This enzyme appears to be a better candidate than AfaA for activating C12 fatty acids for integration into Tambjamine YP1 as it is contained within the *tam* operon. It would also retain the activated fatty acid covalently bound to the enzyme instead of releasing it as a small molecule. This would siphon C12 fatty acids out of the primary fatty acid pool, directly into the Tam pathway instead of allowing C12-CoA to be metabolized by other enzymes.

4.2.2 Cloning, Purification and Characterization

The *tamA* gene was cloned from *P. tunicata* D2 genomic DNA into the pEHISTEV vector to express the protein with a TEV-cleavable N-terminal 6xHis tag (sequence in Appendix 21). This gene was expressed in 2 L of BL21 (DE3) *E. coli* cells at 16 °C for ~18 h with 0.1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). TamA was purified from the cell lysate (Figure 4.4) by nickel immobilized metal affinity chromatography (IMAC). The 6xHis tag was subsequently cleaved by tobacco etch virus (TEV) protease, then purified by a further nickel IMAC and gel filtration chromatography (GFC). The GFC was performed on a HiLoad 16/60 Superdex S200 column, resulting in a peak at 76.4 mL, the expected size of a protein monomer (calibration curve in Appendix 5). This is consistent with other FAAL proteins which are members of the type I adenylation domain superfamily and are generally monomers.[109,110] The final yield of the purified TamA protein was 8 mg/L culture.

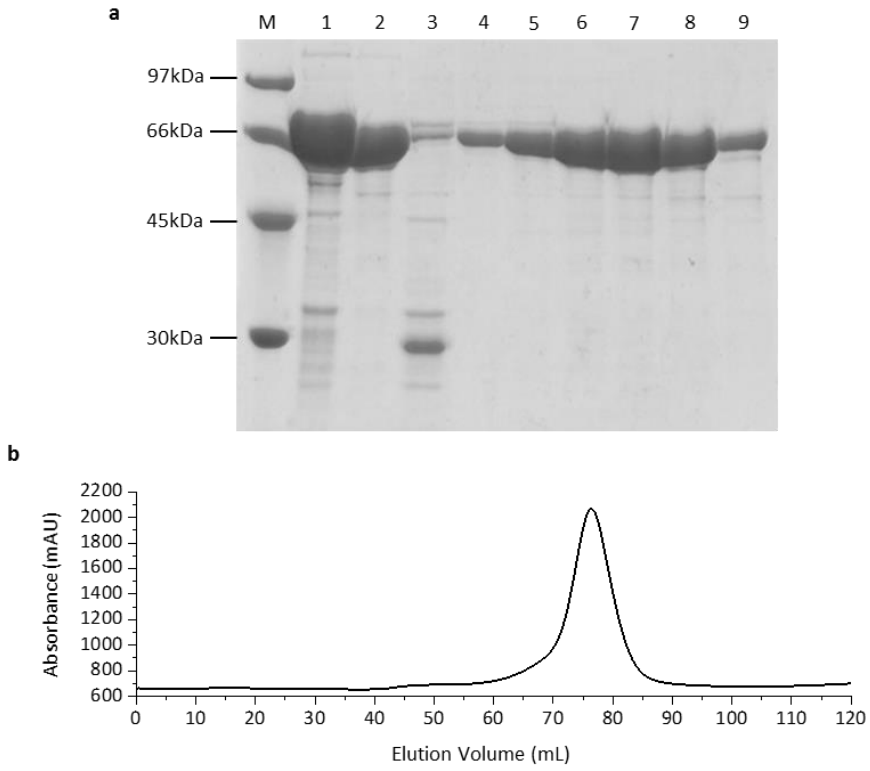


Figure 4.4: TamA purification (**a**) SDS-PAGE analysis: (M) marker, (1) protein after nickel IMAC purification, (2) wash of TEV-cleaved TamA, (3) elution of uncleaved TamA and TEV protease (4-9) 3 mL fractions from the Superdex S200 GFC, (**b**) chromatogram of Superdex S200 GFC monitored at 280 nm, showing peak elution of the protein at 76.4 mL, indicating that the protein is monomeric.

LC ESI-MS analysis (Figure 4.5) showed that TamA, like TamD, was also expressed in its *apo*-form without 4'-PP attached to its ACP. The deconvoluted mass was 75271 ± 2 Da consistent with the calculated mass of 75270 Da.

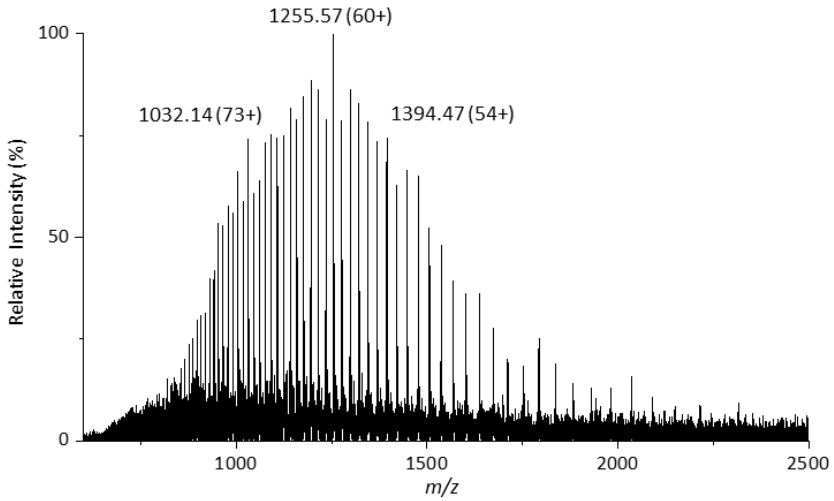


Figure 4.5: Denaturing ESI-MS spectrum of recombinant TamA with a deconvoluted mass of 75271 ± 2 Da consistent with the predicted mass of 75270 Da.

4.2.3 Fatty Acid Selectivity

The promiscuous PPTase Sfp[182] from *B. subtilis* which is commonly employed to install 4'-PP on various ACPs (described in Section 1.4.3) was also used to post-translationally modify TamA. The TamA protein was incubated with Sfp, Mg^{2+} and CoA for ~18 h at 4 °C to attempt the phosphopantetheinylation reaction. Figure 4.6 shows the complete shift of TamA from one mass to another after the 18 h reaction. The mass of 4'-PP is 340 Da but the observed shift was much larger at 513 Da, which was very surprising. The extra 173 Da suggested that the TamA ACP domain was acylated in addition to the 4'-PP modification but it was not clear how this modification occurred.

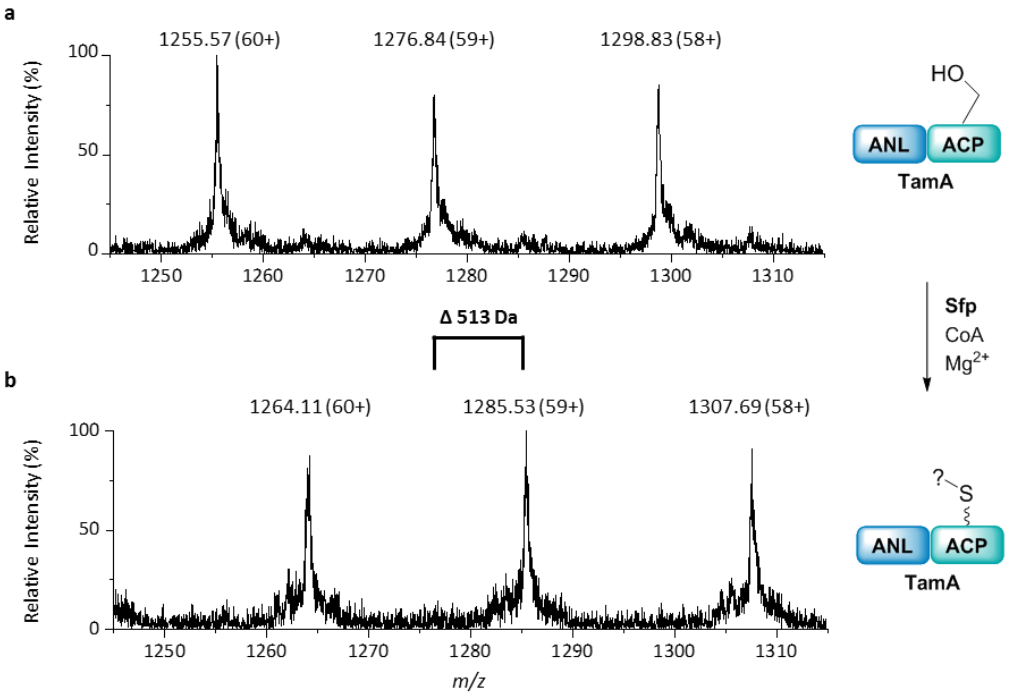


Figure 4.6: Denaturing ESI-MS spectrum of recombinant TamA showing the 60 - 58+ charge states (**a**) as purified, (**b**) after incubation with Sfp, CoA and Mg^{2+} at 4 °C for ~18 h. The mass change is presented as the deconvoluted mass difference after 4'-PP modification of the protein.

For TamA to become acylated in the absence of fatty acid and ATP in the reaction, it was hypothesized that the recombinant TamA could be isolated from the *E. coli* expression host with bound acyl-adenylate. This adenylate would have to be stably bound in the TamA active site throughout the protein purification process and be poised to react with the ACP once it has been 4'-PP modified. The adenylate intermediate would not be visible in the denaturing MS analysis of *apo*-TamA if it was non-covalently bound in the active site.

4.2.4 Non-Covalent Intermediate Binding

Using non-denaturing native mass spectrometry, the as purified TamA was analyzed for binding of an acyl-adenylate, utilizing a method previously employed to identify a BioW-adenylate complex.[111] In this method, the enzyme is infused into the MS intact, so that any non-covalently bound species will remain attached during the analysis. The enzyme was desalted into 100 mM ammonium acetate, pH 8 for direct infusion onto a Waters Synapt

instrument (carried out with the assistance of Dr. Van Kelly). Interestingly, the native MS (Figure 4.7) shows that the as purified protein exists in two forms.

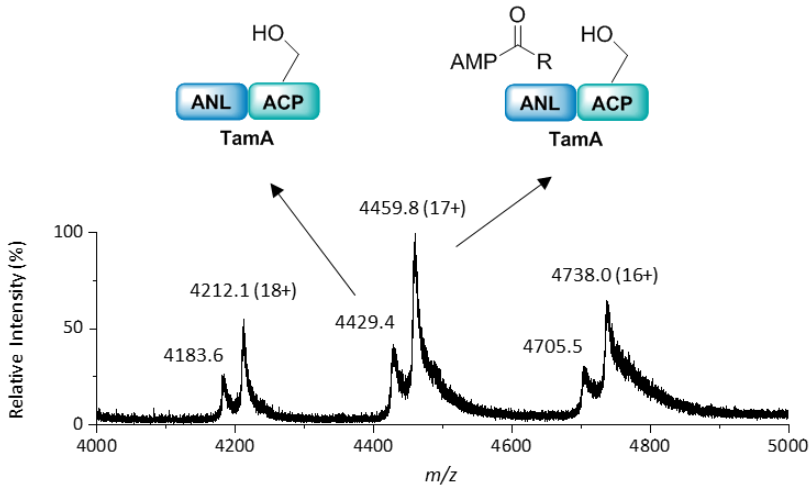


Figure 4.7: Native ESI-MS of recombinant TamA showing three charge states (18+ to 16+) and two species with deconvoluted masses of 75280 Da and 75800 Da respectively, corresponding to *apo*-TamA and acyl-adenylate bound TamA.

The lower abundance species has a deconvoluted mass of around 75280 Da which is consistent with the calculated mass of 75270 Da. There is a larger mass error associated with native mass spectrometry due to peak broadening caused by sodium ions bound to the protein surface.[196] However, the second species (of much higher abundance), has a deconvoluted mass of 75800 Da which is an increase of ~520 Da. C12 fatty acid adenylate has a mass of 529 Da which is within the error of the calculated mass difference. Though this data shows that there is a high likelihood that the protein has an acyl-adenylate non-covalently bound, it is not accurate enough to determine the nature of the acyl chain. The denaturing MS suggested that it is an acyl chain which has a mass of 173 Da however, this mass does not correspond to any unsaturated fatty acid. One explanation for this mass difference is that it corresponds to a mixture of fatty acids which are not resolved due to the large size and high charge state of the protein.

4.2.5 ACP Cloning, Purification and Characterization

To reduce the size of the protein and allow resolution of smaller mass differences, the C-terminal ACP domain was cloned without its ANL partner, so it could be used *in trans* to the full-length protein. The ACP domain is predicted to span residues 582-675 of TamA while the FAAL domain consists of residues 1-572. Therefore, to ensure the full ACP domain was encompassed, residues 578-675 were cloned to be used as a stand-alone ACP (Appendix 22). This protein was also expressed in the pEHISTEV vector for ~18 h at 16 °C with 0.1 mM IPTG in 2L of BL21 (DE3) cells.

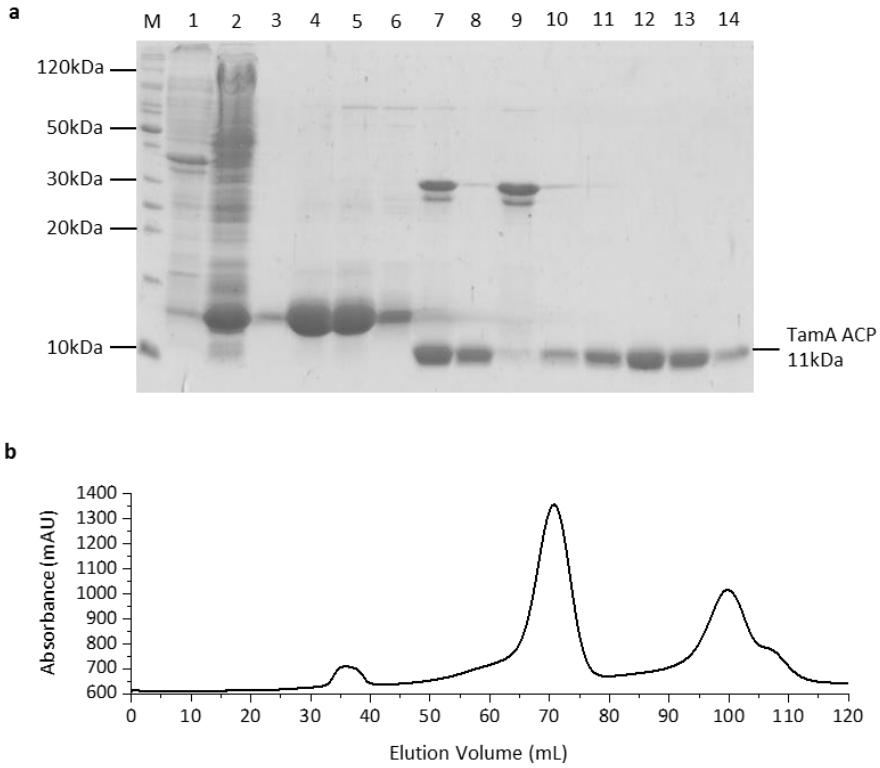


Figure 4.8: TamA ACP purification (**a**) SDS-PAGE analysis: (M) marker, (1) insoluble and (2) soluble fractions of the cell pellet, (3-6) 3 mL fractions from Nickel IMAC purification, (7) protein after TEV cleavage, (8) wash of TEV-cleaved TamA ACP, (9) elution of uncleaved TamA ACP and TEV protease (10-14) 3 mL fractions from the Superdex S75 GFC, (**b**) chromatogram of Superdex S75 GFC monitored at 280 nm showing peak elution at 70.8 mL, consistent with a protein monomer.

The protein was purified by the same method as full-length TamA (Figure 4.8). The TamA ACP GFC produced a symmetrical peak on a HiLoad 16/60 Superdex S75 gel filtration column at the expected elution volume (70.8 mL) for the monomer with a yield of 4 mg/L culture. LC

ESI-MS analysis of purified TamA ACP, like TamA, showed the protein to be in its *apo*-form without 4'-PP (predicted mass: 10967.5 Da, observed mass: 10967.1 ± 0.2 Da, Figure 4.9).

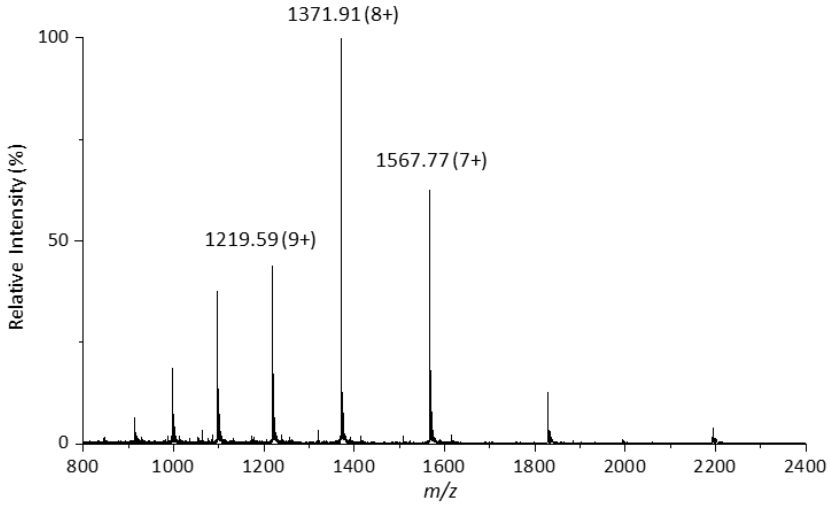


Figure 4.9: Denaturing ESI-MS spectrum of recombinant TamA ACP with a deconvoluted mass of 10967.1 ± 0.2 Da consistent with the calculated mass of 10967.5 Da.

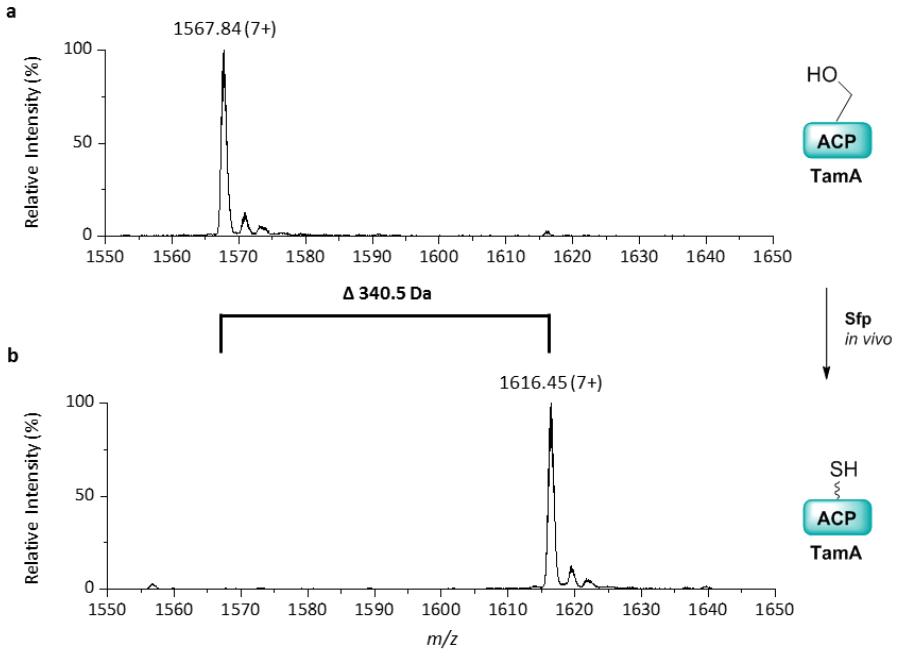


Figure 4.10: Denaturing ESI-MS spectrum of recombinant TamA ACP showing the 7+ charge state (**a**) as purified and (**b**) after coexpression with Sfp. The mass difference presented is the deconvoluted mass change after 4'-PP modification.

To avoid performing the 4'-PP modification reaction using Sfp *in vitro,* the two proteins were coexpressed in BL21 (DE3). An untagged construct of Sfp was utilized to allow separation of the two proteins at the first purification step. After coexpression, full conversion of TamA ACP to the *holo*-form was observed at a mass of 11307.6 ± 0.2 Da (predicted mass: 11307.8 Da, Figure 4.10). Unlike the full-length TamA enzyme, the ACP mass change was correct for the 4'-PP modification on its own (observed: Δ 340.5 Da, predicted Δ 340.1 Da).

4.2.6 Fatty Acid Specificity

The *holo*-TamA ACP was incubated 1:1 with as purified *apo*-TamA to attempt transfer of the fatty acid adenylate from *apo*-TamA to the *holo*-TamA ACP domain (reaction scheme in Figure 4.11a). If the reaction is successful then the acyl chains attached to the stand-alone ACP domain will be detected with a greater accuracy.

When the LC ESI-MS was repeated after incubation, gratifyingly, two new species were observed in the TamA ACP spectrum (Figure 4.11c). The new peaks are 168.1 Da and 182.1 Da larger than *holo*-ACP. These masses are within the error for saturated C11 (168.3 Da) and C12 (182.3 Da) acyl chains. As the two acyl chains are very similar in size, they can account for the single unresolved peak observed in the full-length TamA spectrum suggesting a mass increase of 173 Da.

This suggests that TamA is highly selective as it only adenylates two fatty acid chain lengths from the large fatty acid pool in *E. coli.*[63,170] Though we would expect only C12 to be selected because this is the reported tail length of tambjamine YP1, tambjamines with varying tail lengths have been observed in the same organisms.[28,197]

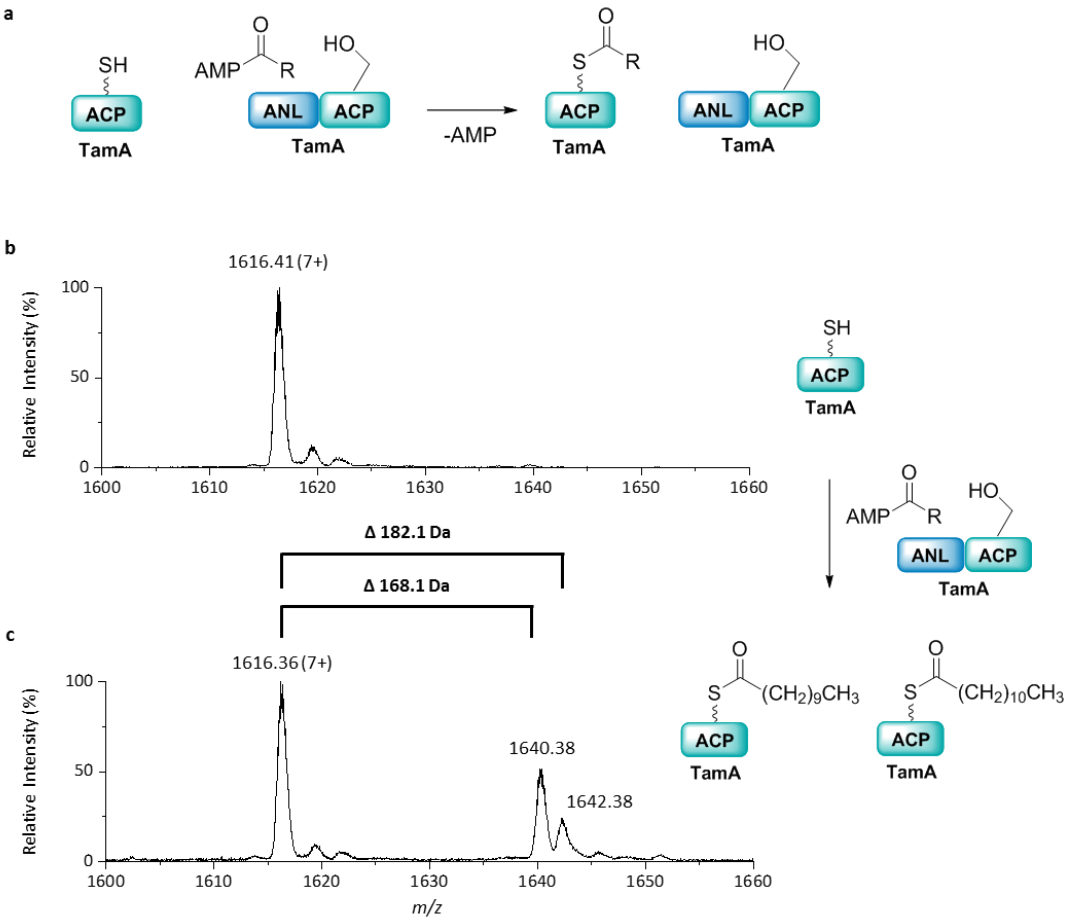


Figure 4.11: (**a**) Reaction scheme showing the TamA ACP domain reacting with as purified TamA (R = unknown hydrocarbon). Denaturing ESI-MS spectrum of *holo*-TamA ACP showing the 7+ charge state (**b**) as purified and (**c**) after incubation with as purified TamA. The mass differences are presented as deconvoluted mass changes.

This new *in trans* assay was further exploited to explore the specificity of TamA for fatty acid chain length. However, instead of using the *apo*-TamA as purified, the full-length protein was modified using the previously described Sfp reaction. This causes any fatty acids bound in the TamA active site after purification to be transferred onto the attached ACP. Therefore, the active site is now likely to be empty and able to adenylate other fatty acids in the reaction (scheme in Figure 4.12). Therefore, 10 µM acyl-TamA, along with 10 µM stand-alone *holo*-TamA ACP were incubated with 1 mM fatty acid of different chain lengths (C2-C16, from a 10 mM stock in dimethyl sulfoxide, DMSO), 5 mM ATP and 10 mM $MgCl_2$. The reaction was then subjected to LC ESI-MS to detect transfer of the fatty acids onto the stand-alone ACP catalyzed by the full-length protein (Figure 4.12).

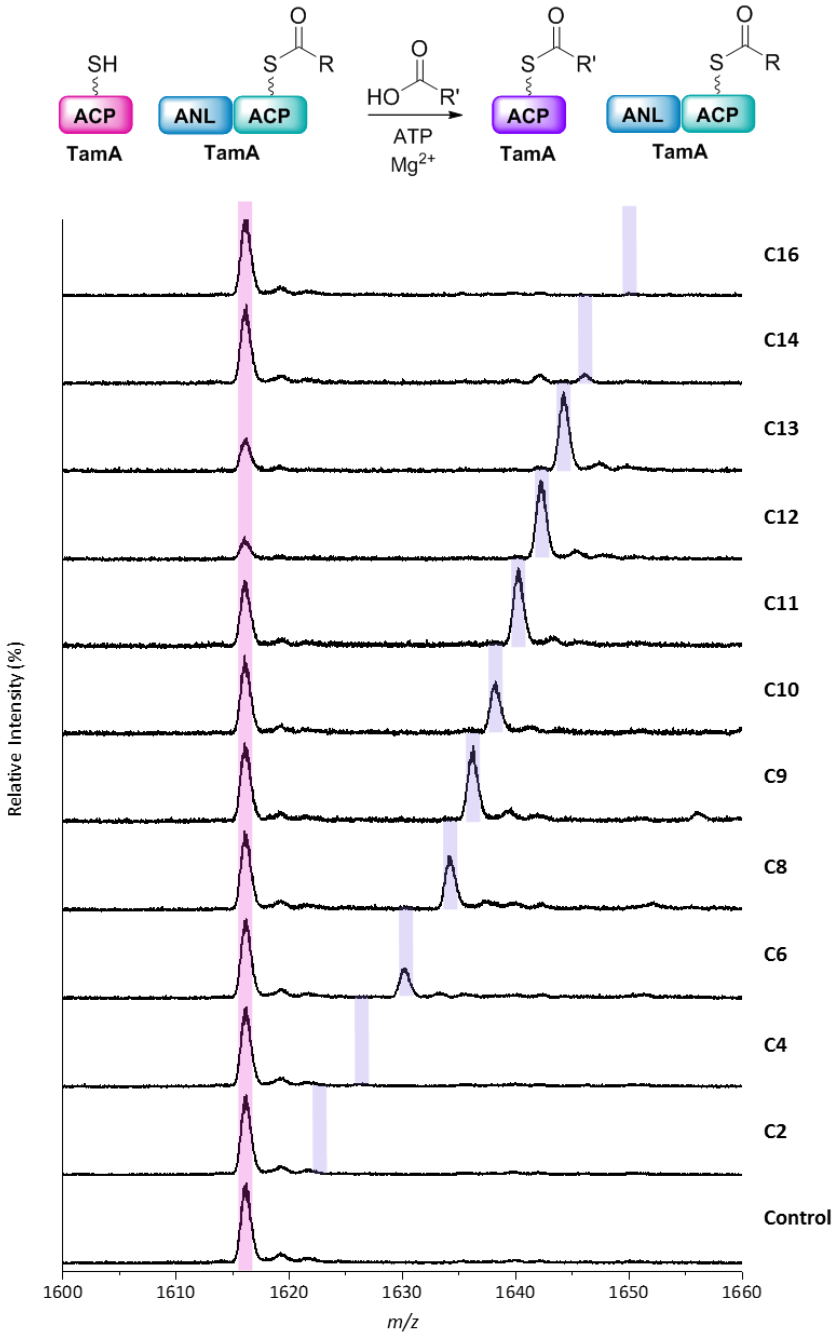


Figure 4.12: Denaturing ESI-MS of TamA ACP after incubation with acyl-TamA, Mg^{2+}, ATP and fatty acids of chain lengths varying from C2 to C16. *Holo*-TamA ACP peaks are highlighted in pink with the expected acyl-TamA ACP peaks highlighted in purple. The data show that TamA can accept fatty acid chain lengths from C6-C13.

The LC ESI-MS data highlights TamA's substrate promiscuity and shows that it can accept fatty acids between C6 and C13 in length. The C14-ACP is barely detectable in contrast to the abundance of C12 and C13 suggesting the protein has a defined hydrophobic cavity for the

fatty acid. The cavity may act as a molecular ruler so extension of the chain length beyond C13 causes efficiency to drop off. The fatty acid producing the largest product peak is C12, further supporting the hypothesis that this enzyme activates C12 fatty acid for the biosynthesis of the tambjamine YP1 alkyl amine tail.

4.2.7 Structural Studies

In order to get a better understanding of the mechanism and selectivity of TamA, the enzyme was subjected to structural studies. The crystallography was carried out in collaboration with Dr. Stephen McMahon in Prof. Jim Naismith's group in the University of St. Andrews. TamA was concentrated to 12 mg/mL after purification and kept at 4 °C or on ice. Sitting drop crystal trials were prepared by Gryphon robot with four bespoke 96-well screens: SA1-4. Two hits for TamA crystals were identified from these screens, wells C1 and F9 in SA3. The best condition was C1 (16% (w/v) PEG3350, 100 mM bicine pH 8.5, 0.12 M $CaCl_2$, 2% hexane diol) which produced crystals that diffracted to 7.5 Å.

However, since there were few hits and many of the conditions screened did not produce crystals or precipitate, this suggested the protein concentration was too low, so the screening was repeated. This time TamA was concentrated to 21 mg/mL. More sitting drop crystal trials were set up with six different 96-well screens: JCSG+, WIZARD and PEGS1-4. The best conditions whose crystals were shot on the St. Andrews X-ray source are summarized in Table 4.1. The crystals from JCSG+, well G11 were the only crystals that diffracted to 6 Å. Though this was better diffraction than the previous screen, this is still low resolution and crystals diffracting to ~2.7 Å are necessary for molecular understanding of the structure.[198] These crystallization conditions were optimized by varying the BIS-Tris pH from 5.5 to 6.5 and ammonium sulphate concentration from 1.75 to 2.25 M. The optimisation was set up by hand in a large 6x4 plate with hanging drops. The same concentration of protein as the screening (21 mg/mL) was used with three drops containing different ratios of protein: precipitant which were 1:1, 2:1 and 1:2. Though crystals grew in several of these wells, the diffraction of the crystals did not improve with the best diffraction only 8 Å resolution.

Screen	Well	Conditions
WIZARD	C9	100 mM CAPS pH 10.5 2 M ammonium sulfate, 200 mM lithium sulfate
WIZARD	H1	100 mM Tris base pH 7.0, 1 M potassium sodium tartrate, 200 mM lithium sulfate
JCSG+	D11	70 mM sodium acetate pH 4.6, 0.14 M calcium chloride dehydrate, 14% (v/v) 2-propanol, 30 % (v/v) glycerol
JCSG+	E2	100 mM sodium cacodylate pH 6.5, 2 M ammonium sulfate, 200 mM sodium chloride
JCSG+	G11	100 mM BIS-Tris pH 5.5, 2 M ammonium sulfate

Table 4.1: Conditions of TamA crystals that were shot on the University of St. Andrews X-ray source.

Since crystals of TamA were not suitable for X-ray analysis, to get an insight into the enzyme's structural elements, a model of the protein was created based on its sequence and homology to proteins of known 3D structure. Though the adenylation domain of TamA is very similar to FAALs, no structures of FAALs fused to ACPs exist as these enzymes usually act *in trans* to their ACP partners. The structures in the PDB with highest similarity to TamA are the adenylation domains and peptidyl carrier proteins (PCPs) from NRPSs. Six structures of four proteins were used to create a model of the TamA protein structure (*E. coli* EntF: 5T3D, 5JA2;[199,200] *Acinetobacter baumannii* AB3403, 4ZXI;[200] *Brevibacillus parabrevis* LgrA, 5ES8;[200] *Geobacillus sp.* strain Y4.1MC1 SrfA-C, 5U89, 2VSQ,[201,202] alignment in Appendix 23). The model was created with the Phyre2 software in intensive mode (Figure 4.13).[203]

The model highlights the ANL domain in blue and the ACP in green. The first view (Figure 4.13a) clearly depicts the ANL domain with the expected type I adenylation fold (described in Section 1.4.5). This fold contains a large N-terminal and smaller C-terminal domain separated by a hinge region (pink). The hinge allows rotation of the two domains with respect to one another to allow for the conformational change required to carry out adenylation and thiolation respectively. The P-loop (yellow) is conserved in all ATP binding proteins and anchors ATP in the active site, establishing the active site location. A second loop of ~30 amino acids, is the FAAL insertion loop (orange) modelled with very little confidence as it is not present in the structures used to build the model. However, it is predicted to be a flexible loop which in other FAALs interacts with the ANL domain, stabilizing the adenylation conformation.[190] Binding of the cognate ACP domain results in release of this loop and allows the conformational change required for thiolation. Figure 4.13b better illustrates the ACP

showing it has the characteristic four helical bundle with Ser622, the predicted 4'-PP site, (red spheres) sitting at the top of helix 2. This is where the 4'-PP modification is observed in other ACP structures.[63,69] The model also shows the two domains interacting with one another. With this interaction conformation, modification at Ser622 would allow the 4'-PP arm to stretch into the ANL active site cavity, positioning it adjacent to the adenylate ready for thiolation.

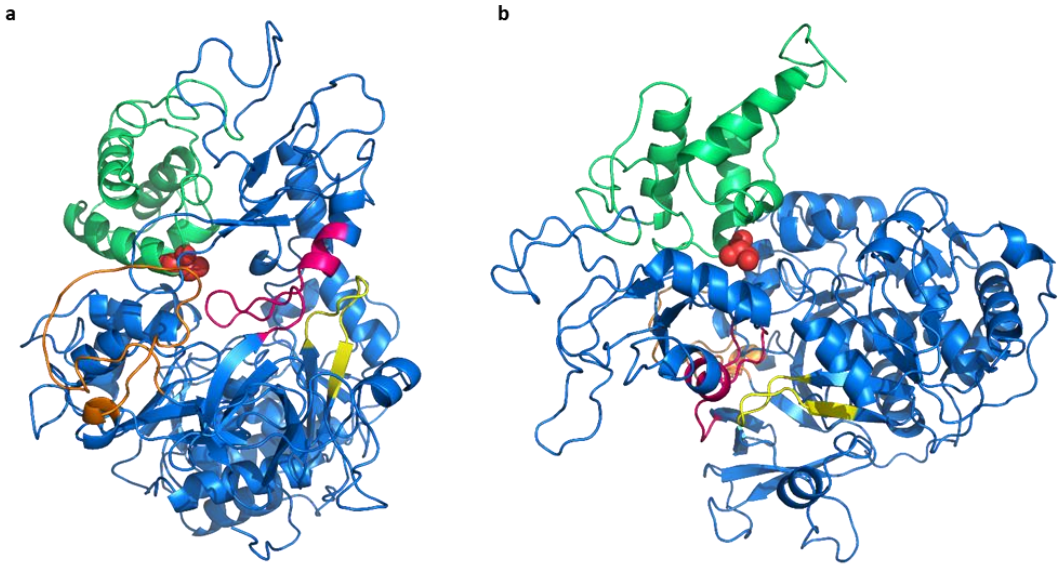


Figure 4.13: Homology model of the TamA structure constructed with Phyre2 from two angles (**a**) and (**b**) showing the ACP domain in green with the 4'-PP modified serine residue Ser622 shown as red spheres, the ANL domain in blue with the hinge region (pink), the ATP-binding P-loop (yellow) and the FAAL insertion loop (orange).

4.2.8 4'-PP Location

According to the bioinformatic analysis and modelling of TamA, the 4'-PP modification is predicted to occur on Ser622. Experimental verification of the position of modification would help to validate the structural model. This is especially important as TamA unusually contains two DSV sequences in the ACP domain, which is the conserved motif for a 4'-PP binding site.[69] To ensure that the second motif, containing Ser622, is the one post-translationally modified, *holo*-TamA was subjected to digestion by trypsin. Digested TamA was infused onto the FT-

ICR MS with the help of Dr. Joanna Simpson (Figure 4.14a). The spectrum shows several peptides with excellent coverage over the protein sequence (98%).

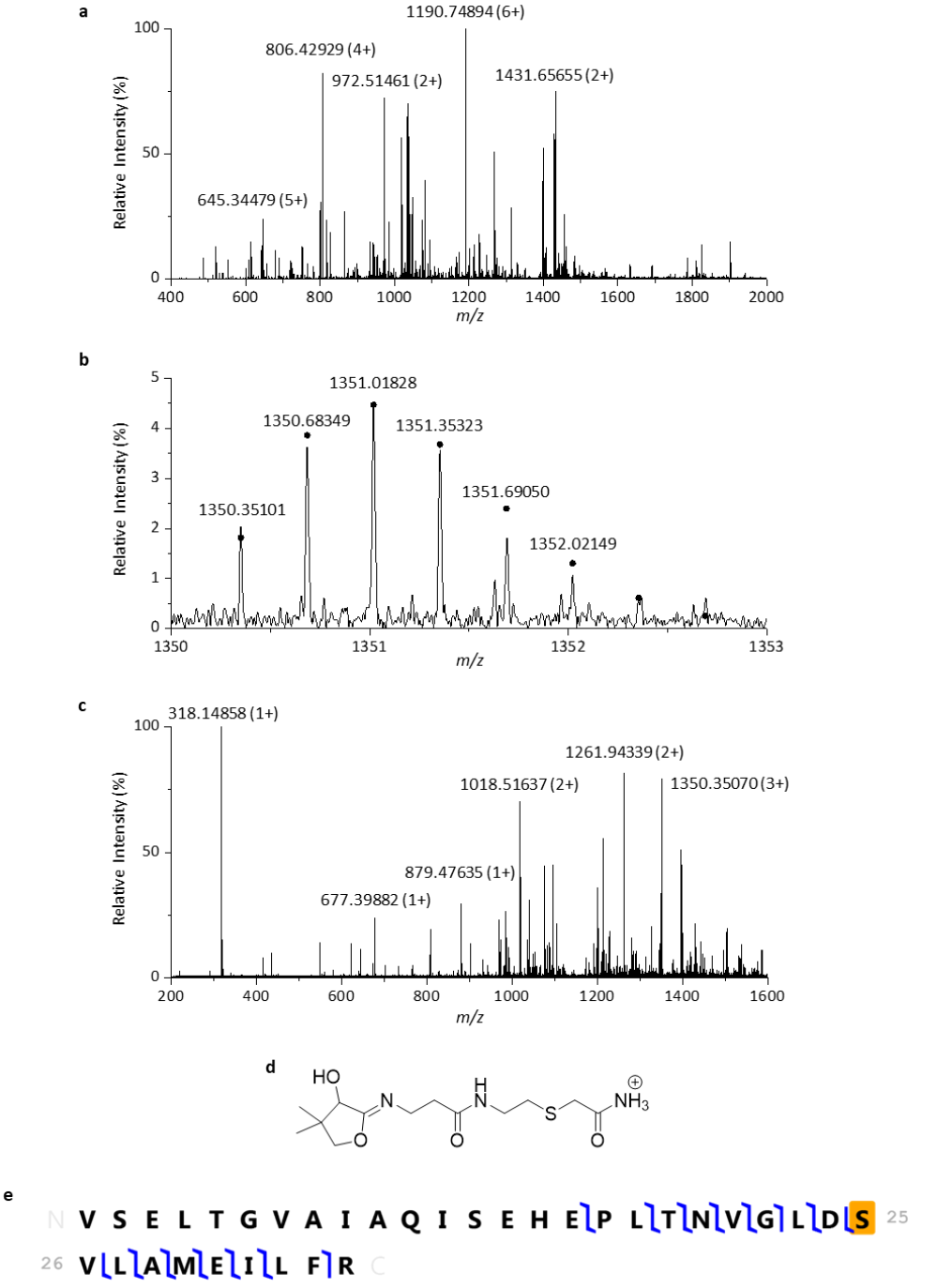


Figure 4.14: FT-ICR mass spectrum of trypsin digested TamA (**a**) full spectrum, (**b**) zoomed in over the 4'-PP modified peptide, (**c**) MS/MS analysis of this peptide, (**d**) structure of the phosphopantetheine ejection ion and (**e**) y- and b-ions of the 4'-PP modified peptide identified in the MS/MS spectrum.

The peptide of interest containing the second DSV sequence with the 4'-PP modification is illustrated in Figure 4.14b. The predicted *m/z* distribution for this peptide is also displayed on the spectrum as dots which fit the experimental data well. The peptide molecular ion was also isolated and subjected to MS/MS analysis to further confirm its identity (Figure 4.14c). The largest peak in the spectrum is the *m/z* consistent with 4'-PP ejection. When 4'-PP modified peptides or proteins are subjected to MS/MS, the 4'-PP is ejected as the molecule shown in Figure 4.14d[204] (predicted $[M+H]^+$: 318.14820 Da, observed: 318.14858 Da). The *m/z* values for the MS/MS were also entered into ProSight Lite software[205] and these attributed to the possible y- and b- fragment ions of the peptide containing 4'-PP (Figure 4.14e). No peptide containing the first DSV motif with 4'-PP was observed which confirms that Ser622 is the position of the 4'-PP post-translational modification.

4.3 TamT Bioinformatic Analysis and Expression

After activation and attachment of the C12 fatty acid to the TamA ACP, three additional steps are required to reach the final amine tail: the introduction of a 3,4-alkene, thioester reduction and transamination (Figure 4.1). Though reduction must occur before transamination, installation of the double bond could occur at any stage. However, it seems likely that the alkene is installed while the fatty acid is ACP bound. Therefore, *P. tunicata* TamT, a predicted FAD dependent dehydrogenase (DH) is expected to be the subsequent enzyme in the pathway (gene and protein sequences in Appendix 24). TamT was cloned into pET28a and pET22b to yield recombinant proteins with N- and C-terminal 6xHis tags respectively. However, upon test expression of these proteins, both constructs appeared completely insoluble with no protein band observed in the soluble fraction (Figure 4.15).

The predicted tambjamine YP1 pathway to the amine tail (Figure 4.1) suggests that this protein installs the double bond prior to reductive amination by TamH. Though this is likely, so that the reaction is carried out while the substrate is ACP bound, it is also possible that this transformation occurs after release of the amine by TamH. As it is not clear at what stage the alkene is installed or whether it is necessary for the activity of the subsequent enzyme, coupling of TamA and TamH activity without TamT was attempted.

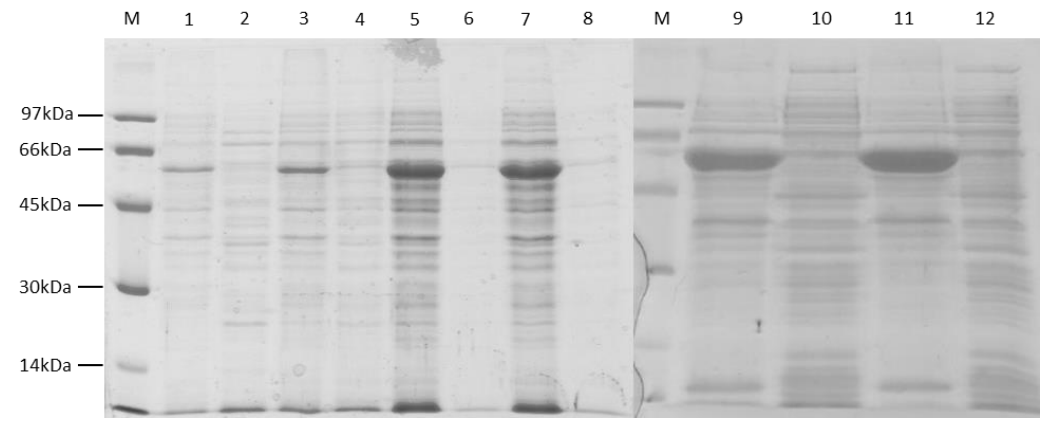


Figure 4.15: SDS-PAGE analysis of test expression of TamT constructs in BL21 (DE3) at 16 °C: (M) Marker (1) insoluble and (2) soluble fractions of expression of TamT pET28a with 0.1 mM IPTG for 3 h, (3) insoluble and (4) soluble fractions of expression of TamT pET28a with 0.5 mM IPTG for 3 h, (5) insoluble and (6) soluble fractions of expression of TamT pET28a with 0.1 mM IPTG for 16 h, (7) insoluble, (8) soluble fractions of expression of TamT pET28a with 0.5 mM IPTG for 16 h, (9) insoluble and (10) soluble fractions of expression of TamT pET22b with 0.1 mM IPTG for 16 h and (11) insoluble and (12) soluble fractions of expression of TamT pET22b with 0.5 mM IPTG for 16 h.

4.4 The TamH Protein

4.4.1 Bioinformatic Analysis

BLAST analysis of the *tamH* gene (sequence in Appendix 25) suggests that the encoded protein is another didomain enzyme whose N-terminal domain is a class III transaminase (TA) and C-terminal domain is a thioester reductase (TR, protein sequence in Figure 4.16). This protein has similarity to PigE from *Serratia sp.* which is involved in the biosynthesis of prodigiosin.[206] PigE is also predicted to contain the same domains as TamH in the opposite orientation with the TR at the N-terminus and the TA at the C-terminus. However, no function for the PigE TR domain has been assigned in the prodigiosin pathway, in fact it appears that only the TA domain is necessary for the biosynthesis. The PigE crystal structure was also solved, to find that the protein had been proteolysed and only the TA domain remained in the crystal.[207] Therefore, no complete characterization of didomain proteins like TamH have been carried out.

The TamH N-terminal class III TA domain (discussed in Section 1.5.3) was aligned with four of its closest homologues in the PDB (*Actinoalloteichus sp.* WH1-2216-6 CrmG, 5DDS;[166] *E. coli* YgjG, 4UOY;[167] *Thermus thermophilus* acetylornithine aminotransferase (AOAT), 1VEF and *Serratia sp.* strain FS14 PigE, 4PPM,[207] Appendix 26) which confirmed the position of the conserved catalytic lysine residue Lys340 and identified other conserved residues involved in PLP stabilization in the active site. These predicted residues include Asp311 which hydrogen bonds with the pyridine nitrogen, Phe206 which engages in a C-H-π interaction with the PLP-ring, Gly114 and Thr370 which stabilize the phosphate group and Gln314 which forms a hydrogen bond with the PLP hydroxyl.

```
MQIRVGQEVL SRESLLESAG LLAQYIRAQG DMLTWQKEDE RIEVLDMVGG FGSTLLGHNH 60
PELLATMQQS LSSLRPMWVQ GAERPVAKQL RNALAQKLLR ETGKKYSIVL LNTGTEAVEA 120
GLKHAQYEFF QRLQHIQQHC DTNWREFKLR LARNEIQLTS EFYLECERLL QQEPIESLEE 180
LQRAVQMRNQ AVFNSSGKIA AFKGDFHGKT RGSLATTYNR DARLPFIANN PDAIFIDDEA 240
QFKATLASWQ KAYFTIEFAP LRLQKKPVNL LTALIYEPIQ GEGGVRPLNA RYCALLNELK 300
LSHPDVAIIA DEIQCGLGRT GQFLESQAIN TPNDYITLAK SLGGGLCKIS ALAVEQTRYH 360
EEFSMLHSTT FADDDLSSAV ALKTLAILER DELTLKAAHL GEQFTTALNA LALEYPDMIA 420
DIRGKGCMLG IELAAQENHP SATIAGLDEQ NMLAMAIAGH LLHRHHIRVL PSLGKRRVLR 480
LQPSAYLDAA NIALVVDALK ETFELIRHHH VATLLAHLIH TDKPHSAFAT AYQHPIREEA 540
IPAGVEKVGF ISHLIDEESL NEIDPSWRLF EGYEQEELNQ HILPITVPGV LARRLVTSAT 600
GRKIELVLYG IQMDAESIEA DRRFNQAKIV RAQVNEAYRL AREEGCRLVG FGGYTSIVTN 660
NCCDYYYNEP ATTSGNALTV AASINTILNS AQDHGINLAK ATVAICGAAG NIGQVHSAIL 720
AKHCHKLLLI TRNVSANNMA MTLNMICEQL YQAVSQDQQG GILVSICRDM LASRIGHEAP 780
KVLIDELKEA LLARQLVRIS EHFNDCQQAD IIVSNTSSPT TVIDAQHVNA HKPVLISDVA 840
VPRDVPVDIV NARPNIKLIR GGVVNLPINP SFTLPGMLLP TGQVYACCGE TMLLGLAGAF 900
SHFSMGALTC EQVEQVQALA AIHGFELNQE KPQYDALQAA S                    941
```

Figure 4.16: The TamH amino acid sequence (UniProtKB: A4C5V8) encoding a 104 kDa protein with a predicted transaminase domain (TA, pink, 56 kDa) and its conserved PLP binding lysine residue (bold) and a thioester reductase (TR, red, 43 kDa) with the conserved GXXGXXG nucleotide binding motif (bold).

The C-terminal domain of TamH is predicted to be a reductase containing a nicotinamide adenine dinucleotide (phosphate) NAD(P)H binding site. These reductases break the thioester bond between an ACP and its acyl chain by attacking the carbonyl with a hydride from NAD(P)H (described in Section 1.4.6). Alignment with TRs of known structure (*Stigmatella aurantiaca* MyxA NRPS, 4U7W;[124] *Mycobacterium tuberculosis* NRPS, 4DQV[122] and *Mycobacterium marinum* carboxylic acid reductase (CAR), 5MSO[125]) identified the

GXXGXXG nucleotide binding motif for NAD(P)H. However, the catalytic triad of threonine, lysine and tyrosine could not be aligned with the TamH TR sequence and there are several gaps in the alignment due to a lack of homology. Therefore, this predicted TR domain is an interesting target for crystallography as very few similar protein structures have been solved.

Based on the TamH bioinformatic analysis and the role determined for TamA in this biosynthetic pathway, TamH is likely to carry out thioester reduction and transamination of a TamA ACP-bound fatty acid (Figure 4.17). A growing number of natural products containing terminal amines have identified the same mechanism of TR and TA cascade to produce this moiety.[121,124] However, none of these exhibit the TamH fused domain structure.

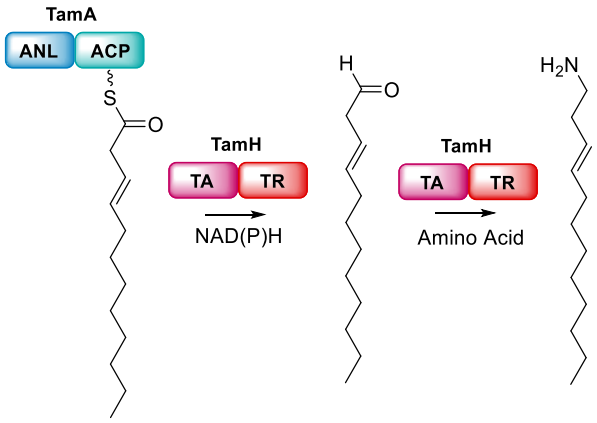


Figure 4.17: Predicted activity of TamH including reduction of acyl-TamA to produce a C12 aldehyde, then transamination of the aldehyde to produce a C12 alkyl amine.

4.4.2 Cloning, Purification and Characterization

In order to study these interesting reactions, TamH was cloned from *P. tunicata* genomic DNA and expressed in pET28a with an N-terminal 6xHis tag. Though expression of this construct produced a large soluble band, the protein appeared to degrade throughout the purification process with two peaks observed in the GFC chromatogram along with other contaminating bands in the SDS-PAGE analysis. Recloning the gene into pEHISTEV, facilitated the introduction of a further purification step by cleavage of the 6xHis tag. However, upon expression of the new construct, a large insoluble band was observed with very little soluble protein. One method of improving protein solubility is to grow the expressing cells in the presence of high concentrations of sorbitol. This places the *E. coli* cells under osmotic stress,

leading to the production of osmotic regulator proteins which promote proper protein folding.[208] Utilizing this expression medium resulted in a much larger soluble protein fraction (Figure 4.18a, lane 2). Therefore, the protein was expressed in BL21 (DE3) in 3 L of this medium and purified from the cell lysate by nickel IMAC. TamH was subsequently cleaved by TEV protease to remove the 6xHis tag, subjected to a second nickel IMAC and finally GFC. The Superdex S200 GFC shows a symmetrical peak with a slight shoulder eluting at 54.6 mL (Figure 4.18). This is consistent with the protein dimer (calibration curve in Appendix 5) which is expected of a PLP protein.[135] The final yield of the protein is quite low at only 2 mg/L culture, however, the protein no longer degrades.

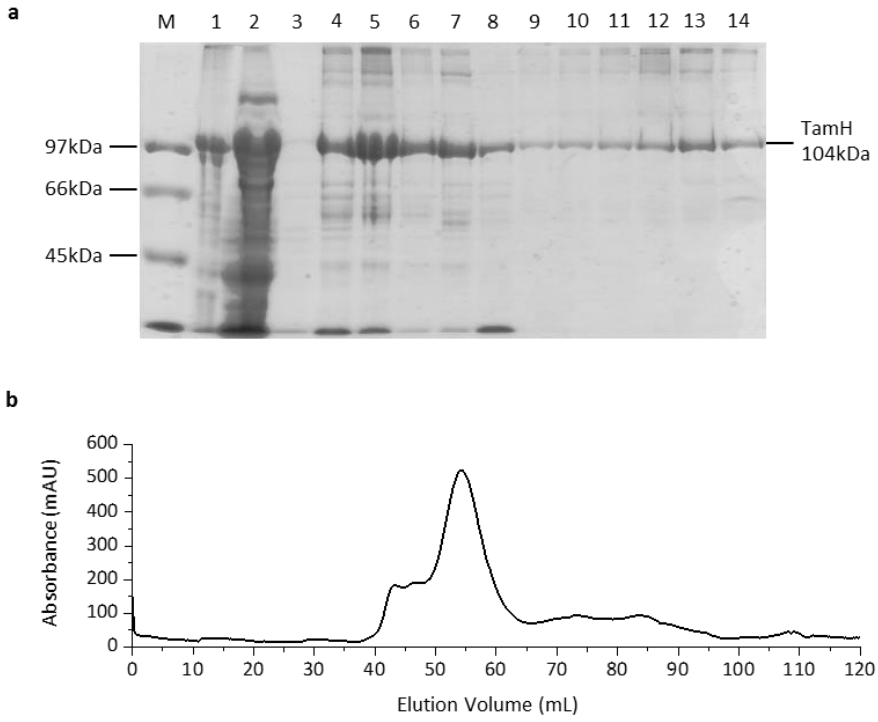


Figure 4.18: TamH purification (**a**) SDS-PAGE analysis: (M) marker, (1) insoluble and (2) soluble fractions of the cell pellet (3-8) 3mL fractions from nickel IMAC purification, (9-14) 3 mL fractions from the Superdex S200 GFC, (**b**) chromatogram of Superdex S200 GFC monitored at 280 nm showing peak elution at 54.6 mL consistent with a protein dimer.

The protein was also subjected to LC ESI-MS however, an accurate mass could not be obtained. This is due in part to the large size of the protein (104 kDa) which is at the limit of detection of the instrument.

4.5 Biosynthesis of C12 Alkyl Amine

Based on the bioinformatic analysis of the TamH sequence and the role that TamA has been shown to undertake, coupling of the two enzymes should result in biosynthesis of C12 alkyl amine from C12 fatty acid (Figure 4.19). This begins with adenylation of the fatty acid and its attachment to the TamA ACP (**A**), followed by TamH reduction to the aldehyde (**B**) and transamination to the amine product (**C**).

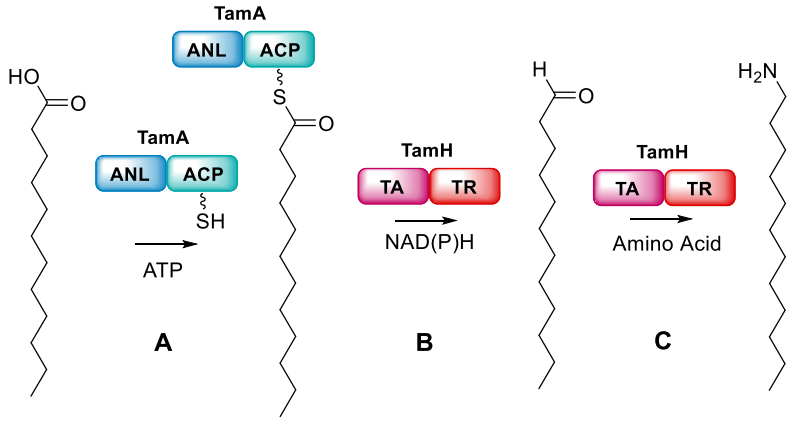


Figure 4.19: Reaction scheme for the biosynthesis of C12 amine from C12 fatty acid (lauric acid) using TamA and TamH. The transformations are (**A**) adenylation of C12 fatty acid and attachment to TamA ACP, (**B**) reduction of C12-TamA ACP to C12 aldehyde with NAD(P)H and (**C**) transamination of the aldehyde with an amino acid to produce C12 amine.

As there are several steps in the pathway to reach the amine product, each transformation was monitored in isolation prior to combination of all the enzymes and substrates. Section 4.2.6 showed that the *holo*-TamA enzyme is able to produce C12-TamA ACP from C12 fatty acid and ATP (**A**), so the TamH reduction of C12-TamA ACP was tested first (**B**). This reaction could be NADH or NADPH dependent, a distinction that cannot be determined from enzyme sequence alone.[121,124] Therefore, separate reactions were set up containing 10 µM TamH, 10 µM TamA, 10 mM $MgCl_2$, 1.25 mM ATP, 1.25 mM C12 fatty acid (from a 25 mM stock in DMSO), 250 µM PLP and 160 µM of either NADH or NADPH. The reaction was monitored for 10 min, taking advantage of the absorbance of NAD(P)H at 340 nm (data not shown). Depletion of the cofactor should result in a decrease in the absorbance at 340 nm as NAD^+ and $NADP^+$ do not absorb at this wavelength. Unfortunately, there was no downward trend in the absorbance, only some fluctuation above and below the baseline suggesting that neither cofactor was depleted. However, since pyridoxal 5'-phosphate (PLP), the TamH TA

cofactor, absorbs in the same range as these molecules (330-420 nm), it could be interfering with the assay, so direct detection of the product was necessary. Aldehydes are difficult to detect by MS because the aldehyde group does not protonate easily. However, the product of the transamination is an amine which ionizes more readily and can be easily detected by LC ESI-MS.

The closest homologue of the TamH TA in the PDB is *Actinoalloteichus sp.* WH1-2216-6 CrmG, a TA involved in caerulomycin biosynthesis.[166] This protein prefers L-glutamic acid as its amino donor and class III transaminases tend to prefer either L-glutamic acid or L-alanine.[133] Therefore, L-glutamic acid was used to test the transamination activity of TamH (**C**) in isolation. TamH was incubated with 5 mM L-glutamic acid and 1 mM C12 aldehyde (from a 10 mM stock in DMSO), at 37°C for ~18 h. Positive and negative controls were also undertaken with the transamination product standard (1 mM C12 amine) and without enzyme respectively.

The extracted ion chromatograms (EICs) of the $[M+H]^+$ *m/z* for C12 amine (186.2222 Da, $C_{12}H_{28}N$) for each of these reactions are presented in Figure 4.20. The EIC for the negative control is flat whereas peaks are displayed in the standard and reaction at the same retention times (12.6 min). These peaks have the same *m/z* of 186.2226 Da which is within 2 ppm error of the calculated mass (186.2222 Da). Though MS usually requires an internal standard to be quantitative,[209] as both the standard and reaction were prepared in the same buffer conditions and run at the same time, the ionization can be assumed to be consistent. Therefore, since the peaks are very close in intensity, this suggests a similar concentration of C12 amine, indicating that the reaction occurred with almost complete conversion. This activity is excellent for an equilibrium TA reaction which often require very high concentrations of amino donors to reach significant conversions.

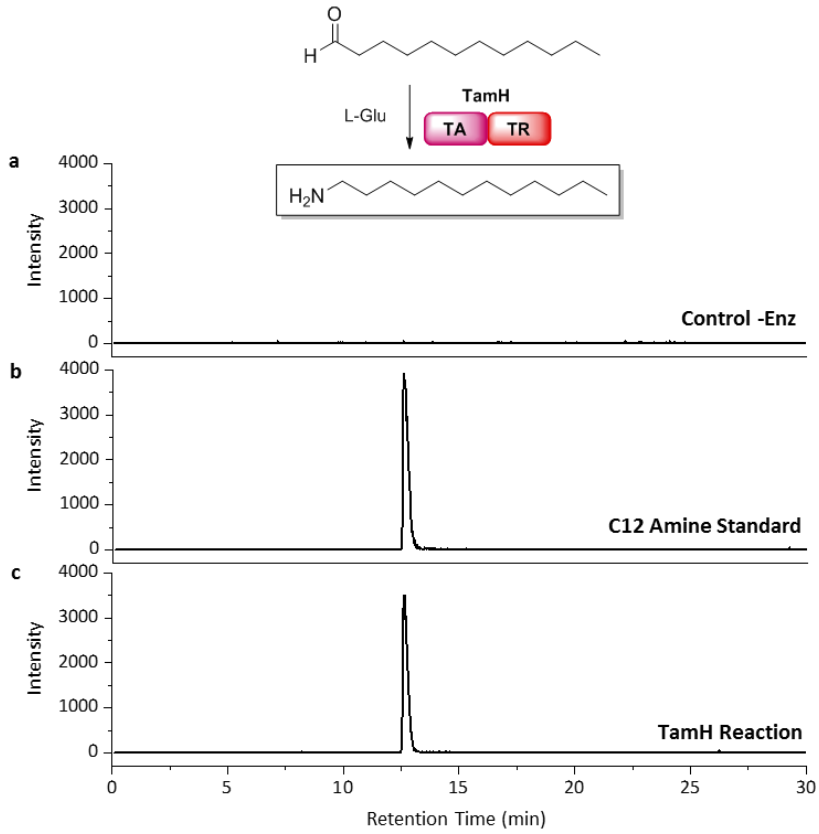


Figure 4.20: Extracted ion chromatograms (EICs) for the C12 amine *m/z* of 186.2222 ± 50 ppm for transamination reactions of TamH (10 µM), L-glutamic acid (5 mM) and C12 aldehyde (1 mM) (**a**) without TamH (**b**) without C12 aldehyde but with C12 amine standard (1 mM) and (**c**) the full reaction. The data show that the reaction produces a peak with the same retention time and intensity as the standard.

Coupling the activities of the two TamH domains with TamA to produce C12 amine (transformations **A**-**C**) from C12 fatty acid was attempted using 10 µM *holo*-TamA, 1 mM C12 fatty acid (from a 10 mM stock in DMSO), 10 mM $MgCl_2$, 5 mM ATP, 10 µM TamH, 2 mM NAD(P)H and 5 mM L-glutamic acid. Reactions were set up with both NADH and NADPH independently. Initially, no product was observed, however a new peak in the chromatogram with an *m/z* of 330.2256 Da appeared. Recent work in the Flitsch group has illustrated that adenylation domains can be used to form amides.[210] This occurs when the ANL domain produces an adenylate intermediate which is intercepted by nucleophilic attack of an amine to form an amide. In the cited work, this occurred because the ACP domain of the CAR enzyme was not 4'-PP modified. However, in the TamA/TamH system the reduction of the ACP thioester may be slow and thus the adenylate is intercepted by the amine of L-glutamic

acid before the 4'-PP thiol can attack. This would result in the production of an L-glutamic acid acyl amide (Figure 4.21, calculated mass: 330.2280 Da for $C_{17}H_{32}NO_5$), explaining the observed *m/z* of 330.2256 Da.

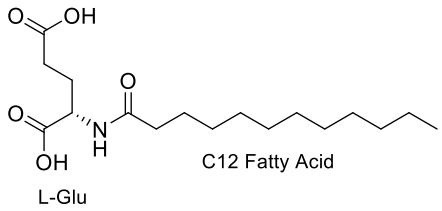


Figure 4.21: Predicted side product for the TamA/H reaction due to reaction of the fatty acid adenylate with L-glutamic acid (L-Glu). Predicted and observed *m/z* for the side product are 330.2280 and 330.2256 Da respectively.

To prevent this side reaction from occurring, the reactants were initially incubated without L-glutamic acid for ~18 h at 37 °C. L-glutamic acid was added the following day, and the reaction incubated for an additional 3 h (scheme in Figure 4.22). EICs show that the control contains no C12 amine and the standard C12 amine peak elutes at 13 min. Excitingly, the same peak is also produced in the full reaction containing the NADH cofactor. This definitively shows that the TamA/TamH reaction system can produce C12 amine from C12 fatty acid using NADH. In contrast, no amine is produced in the NADPH reaction illustrating that TamH is specific for NADH. The reaction was also carried out with 1 mM C12-CoA in place of *holo*-TamA with ATP and fatty acid. Again, no C12 amine was produced (Figure 4.22e), further confirming that this pathway requires the TamA ACP-bound substrate for activity and cannot utilize the free C12-CoA substrate as previously suggested.

Though the MS analysis is not quantitative, the amine reaction product is produced with a lower intensity than the standard, suggesting that the reaction does not go to completion. However, some protein precipitation was observed which may affect the conversion rate. Perhaps this occurred due to the addition of DMSO to solubilize the fatty acid, therefore, decreasing DMSO content or lowering the reaction temperature may improve the conversion. Additionally, some reductases have shown around 10 fold increase in activity with the addition of 200 mM KCl to the reaction.[126,127] It appears the potassium ions may play a structural role in the enzyme. Though there is room for improvement of the reaction system, this preliminary data has provided evidence for the novel role of TamA and TamH in

the biosynthesis of tambjamine YP1 and has exposed an interesting biosynthetic route from fatty acid to alkyl amine.

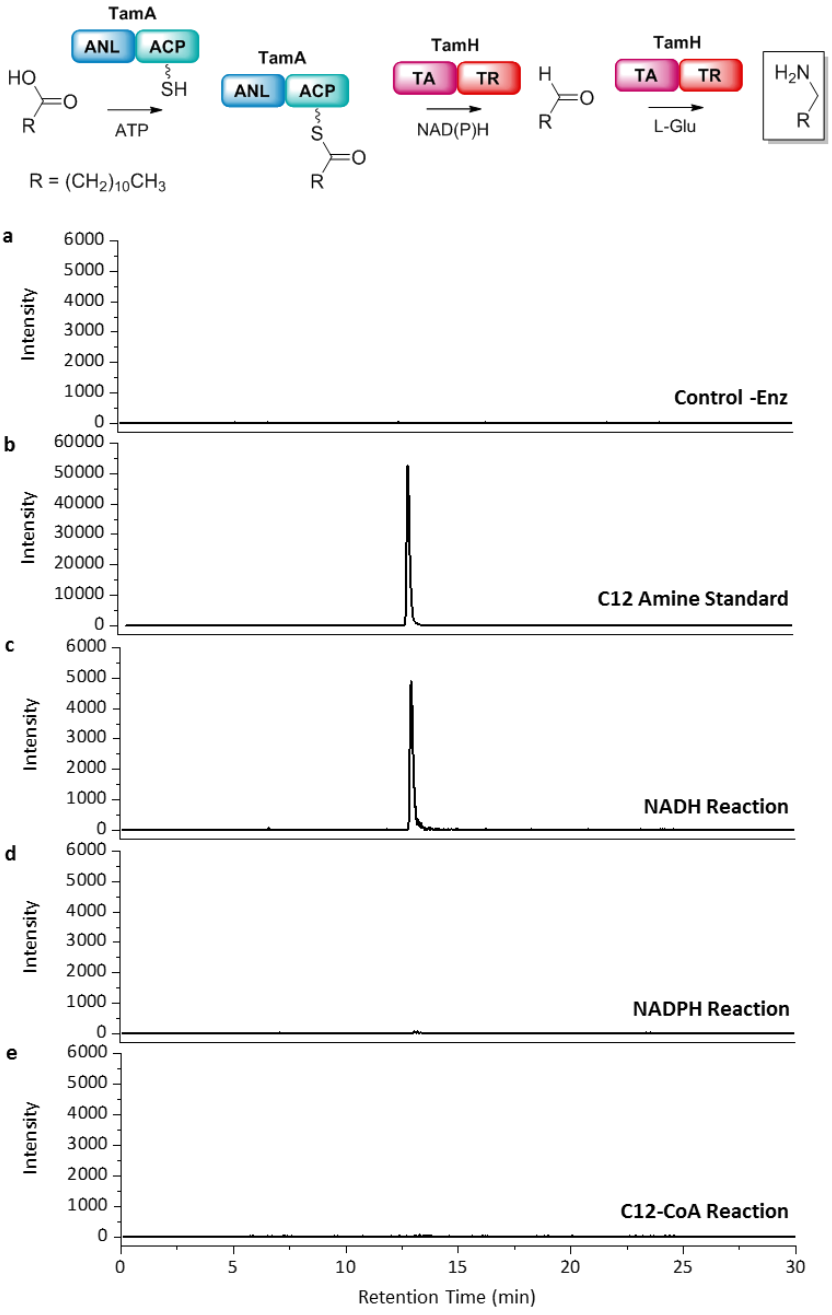


Figure 4.22: Extracted ion chromatograms for the *m/z* of C12 amine (186.222 +/- 20ppm) for the reaction detailed in the scheme containing TamA, C12 fatty acid, ATP, Mg^{2+}, TamH, NAD(P)H incubated at 37 °C for ~18 h before the addition of L-glutamic acid and further incubation for 3 h. EIC for (**a**) the reaction mixture without TamH enzyme, (**b**) the reaction without C12 fatty acid but with C12 amine standard, (**c**) the reaction with NADH, (**d**) the reaction with NADPH and (**e**) the reaction with C12-CoA instead of TamA, ATP and fatty acid.

4.6 Conclusions

The *tam* biosynthetic pathway to the amine tail of Tambjamine YP1 has been successfully probed *in vitro.* The ANL protein encoded by the *tamA* gene which had no previously hypothesized function has been shown to select and activate C11 and C12 fatty acids from the endogenous *E. coli* fatty acid pool during expression. These AMP-activated nucleophiles are ready for transfer onto the TamA ACP domain once it becomes 4'-PP modified. This reaction can also be carried out *in trans* between the full-length TamA and the stand-alone TamA ACP domain. This novel transthioesterification reaction was utilized to screen a panel of fatty acid substrates which showed that TamA accepts fatty acids with chain lengths between C6-C13. It also indicated that C12 fatty acid is the best accepted substrate for TamA, providing evidence for the hypothesized fatty acid activation role of this enzyme in the biosynthetic pathway.

The downstream didomain enzyme TamH, is able to reduce the C12-TamA thioester to produce C12 aldehyde using NADH. It subsequently transaminates the aldehyde using L-glutamic acid to yield C12 amine. The fact that these enzymes work together and that TamH is unable to produce C12 amine from C12-CoA is strong evidence that TamA is indeed the fatty acid activating enzyme in this branch of the tambjamine YP1 biosynthetic pathway, contrary to the previously proposed biosynthetic route.

Full reconstitution of the pathway requires the soluble expression of TamT to confirm its role in installing the 3,4 alkene in the fatty acid and to determine at which stage this modification occurs. Further work on the TamA/TamH system is necessary to improve the turnover of the enzymes as well as determine kinetic parameters and percentage conversion. This would also allow probing of the substrate scope of the system, potentially leading to the production of useful alkyl amines.

5 Amide Biocatalysis

5.1 Introduction

Biocatalysis employs enzymatic catalysts to produce high value chemical compounds.[211] Enzymes are increasingly being used over traditional chemical catalysis because they synthesize products with high stereo- and enantioselectivity in an ecofriendly manner. Natural product enzymes are often chosen as biocatalysts because these proteins carry out interesting and very specific chemistry. Recently, the adenylation (ANL) domain of a carboxylic acid reductase (CAR) was used to produce valuable amides by Flitsch and coworkers.[210] This was achieved by mutating the ACP domain, so it could not be post-translationally modified, leaving the adenylate bound in the ANL active site. Therefore, this adenylate could be intercepted by the addition of an amine, synthesizing an amide, a process eventually used to produce the anticonvulsant drug ilepcimide.

In the previous chapter (Section 4.5) TamA was also shown to produce the amide product dodecyl-L-glutamic acid (Figure 4.21) as a side reaction during the attempt to produce C12 amine. As the ANL domain can accept L-glutamic acid, a fairly bulky substrate, in the amide forming reaction, it may be able to form other interesting and relevant amides according to the reaction scheme in Figure 5.1.

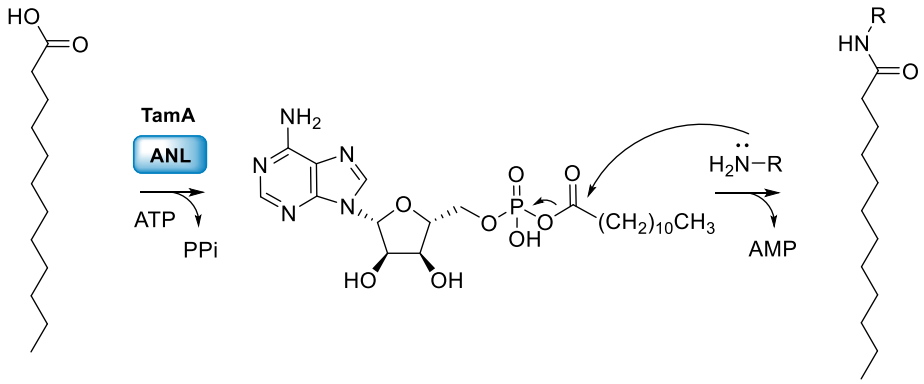


Figure 5.1: Reaction scheme of the TamA ANL domain producing C12 amides by attack of a free amine on its C12-adenylate to generate the amide.

The amide synthesizing reaction does not require the ACP domain but only requires the adenylation activity of the protein. In order to study the action of this enzyme domain on its own, the ANL domain was cloned without the TamA ACP. The substrate specificity of the

amide synthase reaction was probed for both the fatty acid substrate and the amine substrate in order to produce interesting amide products.

5.2 TamA ANL Cloning, Purification and Characterization

Rather than mutating Ser622 in TamA to remove the ACP 4'-PP binding site (as previously reported),[210] the ACP domain was deleted from the adenylation domain completely. Several constructs of differing lengths in pEHISTEV were produced (576, 581, 586 and 591 residues) as the FAAL domain is predicted to span residues 1-571. The constructs were test expressed in BL21 (DE3) (Figure 5.2), all of them showing protein of the expected size (65 kDa). Protein length did not have an effect on the solubility of this protein so the shortest construct (TamA ANL_{576}, sequence in Appendix 28) was used for subsequent experiments.

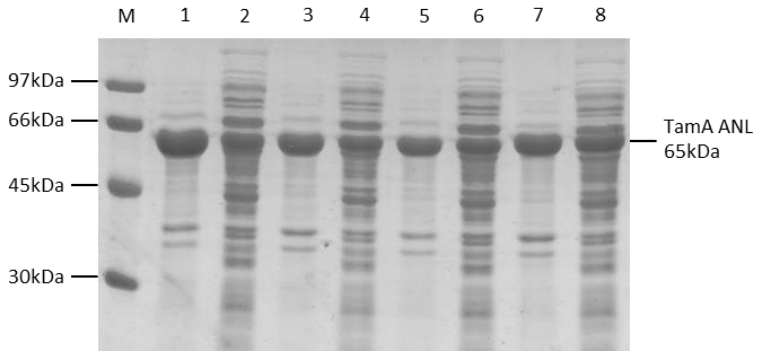


Figure 5.2: SDS-PAGE analysis of test expression of TamA ANL domain constructs (M) Marker (1) insoluble and (2) soluble fractions of expression of TamA ANL_{576}, (3) insoluble and (4) soluble fractions of expression of TamA ANL_{581}, (5) insoluble and (6) soluble fractions of expression of TamA ANL_{586}, (7) insoluble and (8) soluble fractions of expression of TamA ANL_{591}.

Since the ANL domain was expressed in pEHISTEV in the same way as TamA, the same purification method was used, consisting of: immobilized metal affinity chromatography (IMAC), tobacco etch virus (TEV) cleavage, a second IMAC and gel filtration chromatography (GFC, Figure 5.3). The Superdex S200 gel filtration showed a symmetrical monomeric peak at 76.3 mL and produced protein with no visible contaminating bands on the SDS-PAGE analysis. LC ESI-MS analysis of the protein confirmed the deconvoluted mass as 64653 Da ± 1 Da consistent with the prediction of 64651 Da.

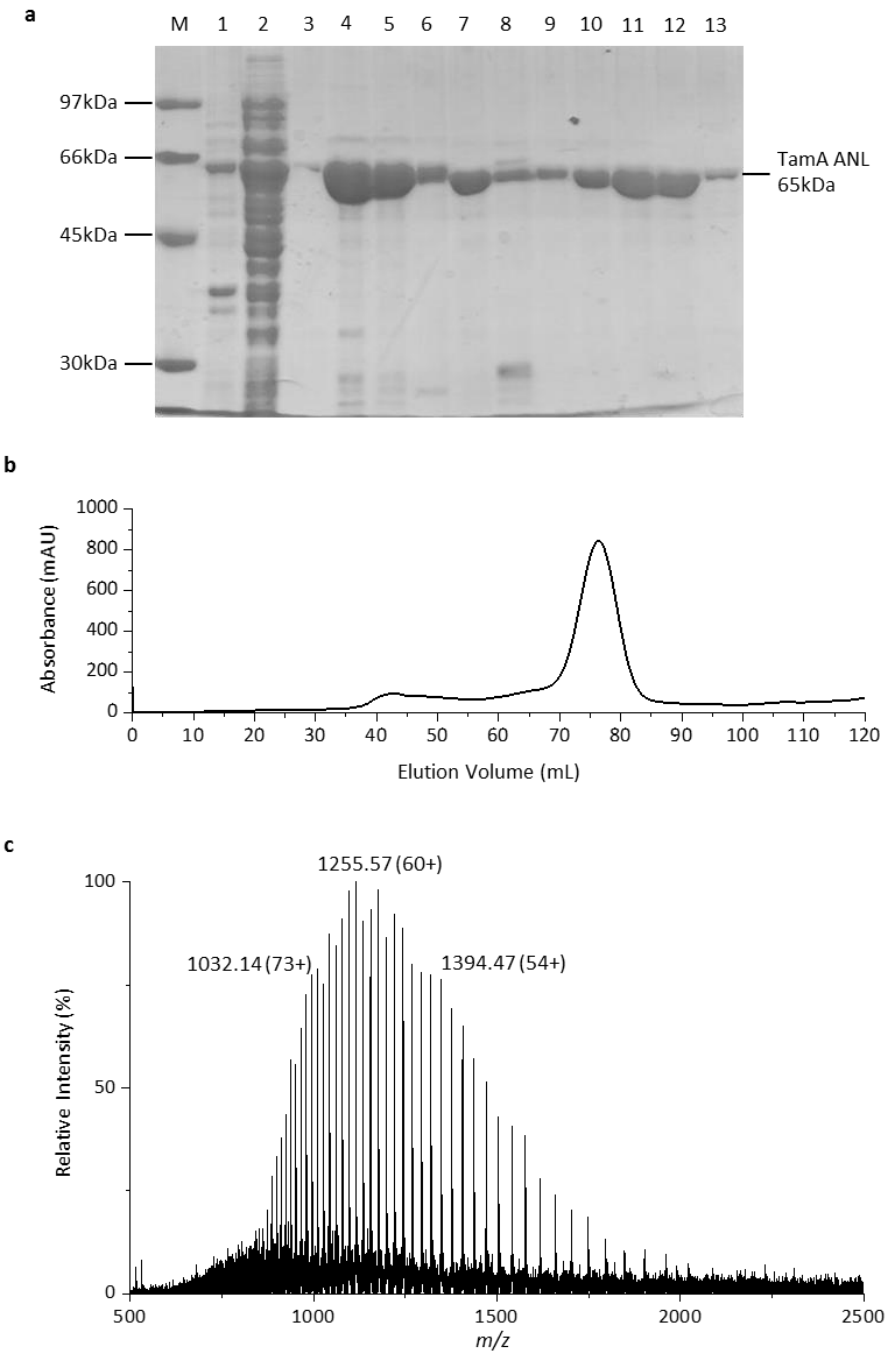


Figure 5.3: TamA ANL_{576} purification (**a**) SDS-PAGE gel showing (M) marker, (1) insoluble and (2) soluble fractions of the cell pellet (3-6) 3 mL fractions from nickel IMAC purification, (7) wash of TEV-cleaved TamA ANL, (8) elution of uncleaved TamA ANL and TEV protease (9-13) 3 mL fractions from the Superdex S200 GFC, (**b**) chromatogram of Superdex S200 GFC monitored at 280 nm showing peak elution at 76.3 mL consistent with the protein monomer (**c**) denaturing ESI-MS spectrum with a deconvoluted mass of 64653 ± 1 Da consistent with the predicted mass of 64651 Da.

5.3 Scope of Amide Synthesis

The fatty acid substrate scope for the ANL amide reaction was tested with addition of ammonia to produce simple amides. Reaction conditions for amide synthesis established by Flitsch and co-workers were employed:[210] 10 µg TamA ANL, 100 mM ammonia (pH 9), 5 mM ATP, 10 mM $MgCl_2$ and 1 mM fatty acid in DMSO (final concentration 4% DMSO). The reaction was carried out with a range of fatty acids varying from C2-C16 in length which were analyzed by LC ESI-MS and the ion chromatograms extracted for the predicted mass of each product (Figure 5.4). The amide products for C9-C14 are clearly visible and increase in retention time with increasing chain length. Below C9 no peaks are observed even though production of C6 and C8 amides would be expected, as these were utilized in the TamA substrate specificity assay (Section 4.2.6). However, it may be that these products do not bind to the C18 column used in this LC ESI-MS method and thus are not observed. The rest of the data displays a similar pattern to the TamA fatty acid specificity assay with the C12 amide showing the best conversion and a drop off in intensity after C13. This suggests that the limiting factor is the production of the adenylate as ammonia can trap any chain length.

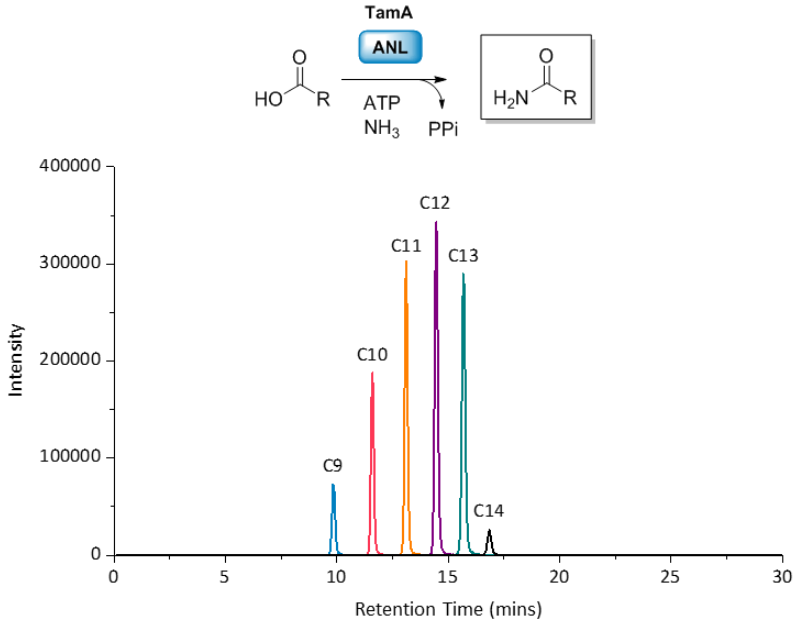


Figure 5.4: Reaction scheme and LC ESI-MS extracted ion chromatograms (EICs) for each of the amide products from the ANL catalyzed reactions of (C2-C16) fatty acids with ATP and ammonia.

Though the TamA ANL is unlikely to catalyze the amine attack on its bound adenylate, the enzyme active site must be large enough to accommodate the entering amine. Therefore, the reaction was attempted with different amine substrates and C12 fatty acid to see if it has any specificity for the amine. L- and D-glutamic acid were used to determine if the enzyme has enantioselectivity since L-glutamic acid was previously accepted (Figure 5.5a,b).

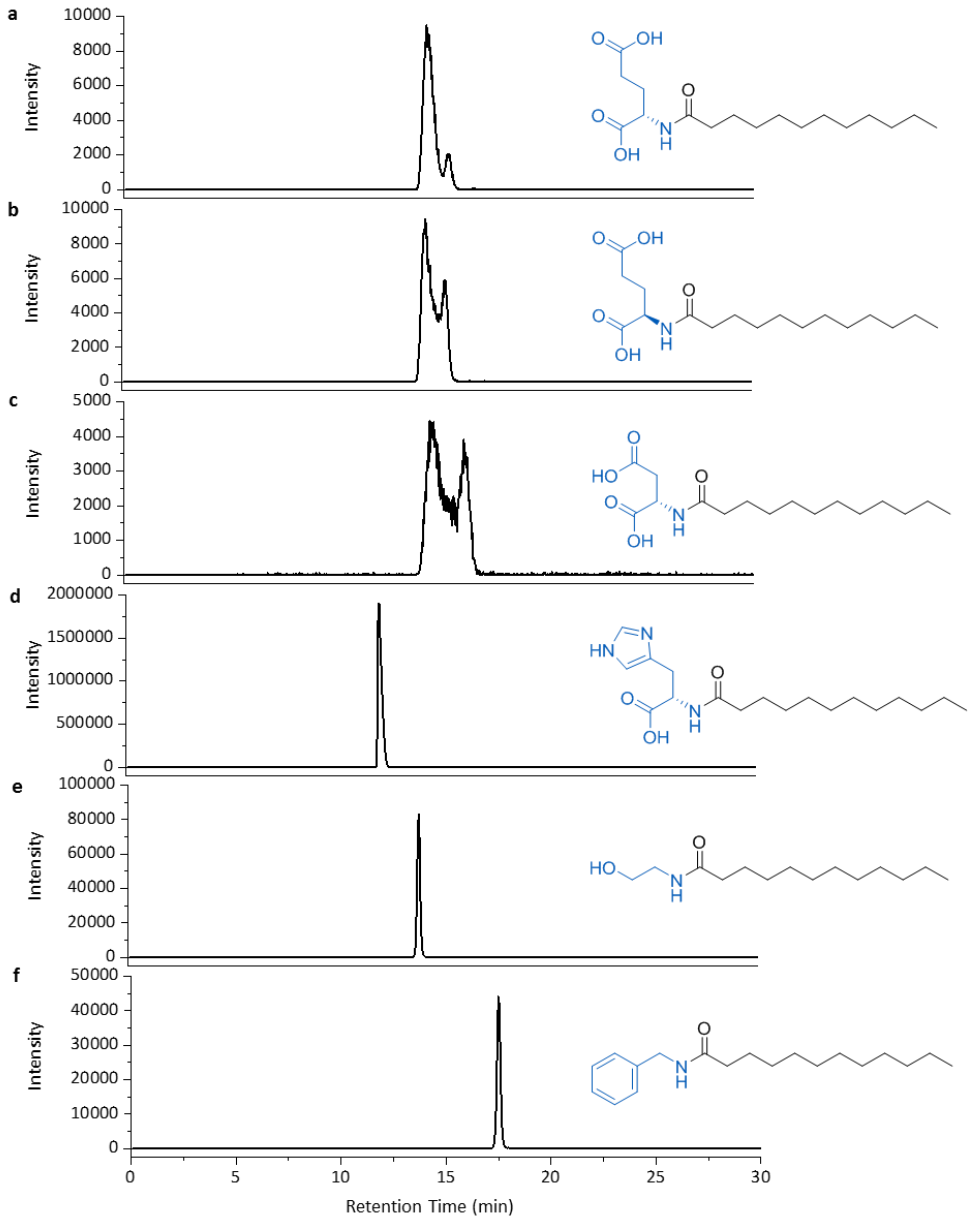


Figure 5.5: LC ESI-MS EICs for the amide products of the TamA ANL catalyzed reactions of C12 fatty acid, ATP, Mg^{2+} with (**a**) L-glutamic acid, (**b**) D-glutamic acid, (**c**) L-aspartic acid, (**d**) L-histidine, (**e**) ethanolamine and (**f**) benzylamine. The data show that all of these substrates are turned over by the TamA ANL.

The enzyme does not appear to have a preference between the two enantiomers as both could be observed with similar intensity in the LC ESI-MS along with the very similar L-aspartic acid which has a side chain one carbon shorter.

Since TamA ANL could be useful biocatalytically, other interesting commercially important targets were identified. *N*-acyl histidines have recently been discovered in *Legionella pneumophila,* a human pathogen.[212] These metabolites are upregulated during the *L. pneumophila* infection cycle and could regulate human signaling pathways, analogously to other identified acyl amides. Since these molecules are important for studying human-pathogen interactions and they are not commercially available, L-histidine was also tested as a substrate for the TamA ANL amide synthesis activity (Figure 5.5d). The LC ESI-MS data clearly shows TamA production of this important molecule.

Palmitoylethanolamide (PEA), the amide produced by the coupling of C16 fatty acid and ethanolamine is a drug which has been widely used to treat chronic pain.[213] This molecule is commercially available from Sigma Aldrich for £72 for 10 mg at the time of writing. However, this simple molecule could be biosynthesized from two cheap starting materials (namely, palmitic acid and ethanolamine). Though the TamA ANL domain does not currently accept palmitic acid, C12 fatty acid could be utilized to produce a similar molecule. Therefore, ethanolamine and C12 fatty acid were combined with TamA ANL to produce C12 ethanolamine (Figure 5.5e). In the future, with the help of a TamA ANL structure, the enzyme could be evolved to accept longer fatty acids and produce PEA.

Finally, a benzylamine acyl amide was also biosynthesized (Figure 5.5f), showing that larger aromatic molecules can also be accepted by the TamA ANL active site. This molecule would be useful for determining the kinetics of this reaction due to its aromatic ring which allows detection of the product by HPLC instead of requiring LC ESI-MS analysis.

5.4 Conclusions

With the growing call for green industrial processes, biocatalysis remains one of the most promising methods of achieving ecofriendly and sustainable reactions. The TamA ANL domain from tambjamine YP1 biosynthesis has proved to be a useful biocatalyst for the production of a variety of interesting fatty acyl amides, with a wide range of applications.

Initial reaction conditions for this transformation have been explored and the substrate specificity probed. Though the full potential of this tool has not yet been determined, this preliminary data illustrates that it can be used to produce exciting *N*-acyl histidines, important molecules in the *L. pneumophila* human-pathogen interaction. Future work will involve determining rates of reaction and conversion to allow scale up of this process to industrial standards. TamA ANL also has the potential to be used to biosynthesize the high value compound PEA and other *N*-acyl amide signaling molecules, if the protein can be engineered to accept longer fatty acids. TamA is an exciting, new biocatalytic tool and further research into the amine substrates that can be employed in this reaction could also open new doors to other useful compounds.

6 Materials and Methods

6.1 Materials and Reagents

Reagents, chemicals and media were purchased from Sigma-Aldrich, Thermo Fisher Scientific or Bio-Rad and used without further purification. Competent cells, pET plasmids and restriction enzymes were purchased from New England Biolabs (NEB) and Life Technologies. Primers were obtained from Sigma-Aldrich and chromatography columns were purchased from GE Healthcare. The *Pseudoalteromonas tunicata* D2 strain was acquired from DSMZ.

6.1.1 Competent *E. coli* Cell Lines

Various *E. coli* cell lines were used to store and amplify plasmids during cloning and for producing recombinant proteins. These are summarized in Table 6.1.

Cell Line	Supplier	Application
C2987	New England Biolabs	High transformation efficiency, plasmid storage and amplification during cloning
DH5α	Life Technologies	Plasmid storage and amplification
BL21 (DE3)	New England Biolabs	Protein expression host

Table 6.1: *E. coli* strains utilized in this work, the supplier and the applications of their use.

6.1.2 Plasmids

Target genes were cloned into various plasmids (Table 6.2) to yield recombinant proteins with specific purification tags. The pEHISTEV plasmid was a gift from Prof. Jim Naismith, University of St. Andrews.[181]

Plasmid	Resistance	Application
pET22b	Ampicillin	Expression of protein with a C-terminal 6xHis tag or no affinity tag
pET28a	Kanamycin	Expression of protein with an N-terminal 6xHis tag

pEHISTEV	Kanamycin	Expression of protein with an N-terminal Tobacco Etch Virus (TEV) protease-cleavable 6xHis tag
pACYC	Chloramphenicol	Expression of protein with no affinity tag

Table 6.2: Expression plasmids utilized in this work, the antibiotic resistance conferred to the host and their applications.

6.1.3 Growth Media

Growth media were prepared by dissolving the components listed in Table 6.3 in distilled water, then sterilizing by autoclaving for 20 min at 120 °C.

Medium	Components
Lysogeny Broth (LB)	10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl
LB Agar	10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl, 15 g/L agar
Super Optimal Broth (SOC)	10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 2% w/v glucose
LB + Sorbitol	LB components were dissolved in 100 mM Na phosphate pH 7.5 along with 0.5 M sorbitol
VNSS Agar[46]	1 g/L peptone, 0.5 g/L yeast extract, 0.5 g/L glucose, 0.01 g/L $FeSO_4 \cdot 7H_2O$, 0.01 g/L Na_2HPO_4, 17.6 g/L NaCl, 1.47 g/L Na_2SO_4, 0.08 g/L $NaHCO_3$, 0.25 g/L KCl, 0.04 g/L KBr, 1.87 g/L $MgCl_2.6H_2O$, 0.41 g/L $CaCl_2$, 0.008 g/L $SrCl_2$, 0.008 g/L H_3BO_3 and 15 g/L agar

Table 6.3: Components of the growth media utilized in this work.

Appropriate antibiotic was added to cool growth media after sterilisation when growing cell lines transformed with plasmid. Kanamycin, ampicillin and chloramphenicol stocks were made with concentrations of 30 mg/mL, 100 mg/mL and 35 mg/mL respectively. The kanamycin and ampicillin stocks were dissolved in water and the chloramphenicol stock in ethanol. The stocks were added to media in a ratio of 1:1000 for final concentrations of 30 µg/mL, 100 µg/mL or 35 µg/mL.

6.1.4 Buffers

Buffers (Table 6.4) were prepared by dissolving components in distilled water and adjusting to the correct pH using concentrated hydrochloric acid or sodium hydroxide. Buffers were filtered through 0.22 µm filters and degassed by sonication prior to chromatography. When pyridoxal 5'-phosphate (PLP) dependent enzymes were purified (TamD, TamD $AOS_{\Delta 96}$, TamH), PLP was added to buffers A-D in a 1:1000 dilution from a stock of 25 mM PLP for a final concentration of 25 µM PLP.

Medium	Components
A	50 mM Tris pH 8, 250 mM NaCl, 10 mM imidazole, 10% glycerol
B	50 mM Tris pH 8, 250 mM NaCl, 500 mM imidazole, 10% glycerol
C	50 mM Tris pH 8, 250 mM NaCl, 1 mM DTT, 10% glycerol
D	50 mM Tris pH 8, 250 mM NaCl, 10% glycerol
E	25 mM Potassium phosphate pH 8, 250 mM NaCl, 1 mM DTT

Table 6.4: Components of buffers utilized in this work.

6.2 Cloning

6.2.1 Overview

Genes were cloned from *P. tunicata* genomic DNA using sequence specific primers containing restriction sites matching the host vector. The primers were used to amplify the gene by polymerase chain reaction (PCR). PCR products and host vectors were digested using two restriction enzymes, then purified by agarose gel electrophoresis. The digested PCR product and vector backbone were ligated to form a new plasmid and transformed into high efficiency competent *E. coli* cells. Single colonies were grown up and the plasmid extracted. To ensure the plasmid harbored the new cloned gene, the plasmid was subjected to analytical digest, which if positive was followed by gene sequencing. Protein subunits were cloned directly from plasmids containing the full gene.

6.2.2 *P. tunicata* genomic DNA

P. tunicata was streaked onto VNSS agar plates and grown at room temperature for 5 days. The cells were scraped into PBS and resuspended. They were subsequently spun down at 17000 xg in a VWR MicroStar 17 centrifuge for microfuge tubes. The genomic DNA was extracted using Qiagen DNeasy Blood and Tissue kit according to the manufacturer's instructions.

6.2.3 Polymerase Chain Reaction (PCR)

Primers were used to amplify specific fragments of *P. tunicata* genomic DNA in a PCR. Primer oligonucleotides were stored as 100 µM stocks at -20 °C after resuspension. These were diluted to 10 µM stocks for use. Primer sequences with restriction sites in bold are presented in Table 6.5.

Primer	Sequence
TamF For	GCCG**CATATG**AGCAAGCAGAAGTATGTGATATCGCTG
TamF Rev	GCGC**CTCGAG**GACTTGTGATCCGCAATAGTCAC
TamD For	**CCATGG**CAGACAATAAAAACACCGCTA
TamD Rev	**CTCGAG**TCATGCCACAAAACCGGGTAATTT
TamD ACP For	CTCT**CCATGG**GAATGACAGACAATAAAAACACCGCTA
TamD ACP_{95} Rev	TTTC**CTCGAG**TCATTGTGCCAGTTTTTGCACTC
TamD ACP_{98} Rev	TTTC**CTCGAG**TCAAAGTCCTAATTGTGCCAGTTTTTG
TamD ACP_{101} Rev	TTTC**CTCGAG**TCATGCTGCTGCAAGTCCTAATTG
TamD ACP_{104} Rev	TTAA**CTCGAG**TCAAGCTTGTGGTGCTGCTGCAA
TamD $AOS_{\Delta 96}$ For	**CATATG**TTAGGACTTGCAGCAGCACCACAAGC
TamD $AOS_{\Delta 96}$ Rev	**CTCGAG**TGCCACAAAACCGGGTAATTTCAC
TamD $AOS_{\Delta 131}$ For	CCGG**CATATG**CCTGACTTATTTAGTAAAGTGCGTAAG
TamD $AOS_{\Delta 128}$ For	CCGG**CATATG**GATGGTAGCCCTGACTTATTTAGTAAAG
TamD $AOS_{\Delta 125}$ For	TTGT**CATATG**TTTGTCGCAGATGGTAGCCCTGACTTAT
TamD $AOS_{\Delta 122}$ For	ATTA**CATATG**CGAGATTTTGTCGCAGATGGTAGCCCT

TamD AOS Rev	CTGT**CTCGAG**TGCCACAAAACCGGGTAATTTCA
TamS For	GG**CCCATGG**TAAGCGAGAATTTGTTAACGC
TamS Rev	**GCGGCCGC**TCATGCATAATTAGGTTCTACATA
TamA For	TCGC**CCATGG**AATGTGAAGCAAGTAGTTTGATT
TamA Rev	GGCC**CTCGAG**CTAACATGATAAATTCAGCTCCT
TamA ACP For	GAGC**CCATGG**CGCCACGTGAAGTGTC
TamA ACP Rev	GGCC**CTCGAG**CTAACATGATAAATTCAGCTCCT
TamA ANL For	TCGC**CCATGG**AATGTGAAGCAAGTAGTTTGATT
TamA ANL_{576} Rev	CGTA**CTCGAG**TCACGCAATTTGCTCGCCTTG
TamA ANL_{581} Rev	GCTA**CTCGAG**TCACGACACTTCACGTGGC
TamA ANL_{586} Rev	CGTA**CTCGAG**TCAATTGCTGATATCAAGCGACAC
TamA ANL_{591} Rev	GATC**CTCGAG**ACTAGCAACACTGTCTTGATTGCTGA
TamT For	GCCG**CATATG**CATGACGATAATACTTTTTATGCTG
TamT Rev Stop	ATAT**GCGGCCGC**TTATATTGGCAAAAGTGAAAACCATTTT
TamT Rev No Stop	ATTA**GCGGCCGC**TATTGGCAAAAGTGAAAACCATTTT
TamH For	ATTA**CCATGG**AAATTCGTGTTGGACAAGAAGTTTTAAGCCGAG
TamH Rev	ATAT**CTCGAG**TCAGCTAGCGGCTTGCAAGGCGTCATATT

Table 6.5: Primer sequences for cloning of target *P. tunicata* genes.

All PCR reactions were prepared on ice with 5x Phusion HF Buffer (10 µL), 10 µM forward primer (1.5 µL), 10 µM reverse primer (1.5 µL), 10 mM dNTPs (1 µL), genomic DNA (~100 ng), Phusion (1 µL), and distilled water to a final volume of 50 µL. Thermal cycling was carried out according to Table 6.6 with annealing temperatures dependent on the melting temperature of the primers used in the reaction. Denaturing and annealing times were dropped to 30 secs for amplification of protein subunits from a plasmid instead of genomic DNA.

Initial Denaturing		98 °C	3 min
Cycle x30	Denaturing	98 °C	1 min
	Annealing	52-58 °C	1 min
	Extension	72 °C	1.5-3 min
Final Extension		72 °C	10 min

Table 6.6: PCR cycling conditions used for amplification of *P. tunicata* genes from genomic DNA.

6.2.4 Gel Electrophoresis and Extraction

DNA gels were prepared with agarose (1 g) in TAE buffer containing 40 mM Tris, 20 mM acetic acid and 1 mM EDTA (100 mL). The mixture was heated until the agarose was dissolved and further cooled to ~50 °C before addition of GelRed (10 µL). The solution was stirred, poured into a casting mold and allowed to set. PCR products and other DNA samples were mixed with 6x Purple DNA loading dye (NEB) and loaded onto the gel next to a marker (Hyperladder I, Bioline). Gels were run at a constant voltage of 100 V for around 50 min. DNA bands were visualized using UV light and DNA bands of interest were excised from the gel using a scalpel. The gel band was weighed and DNA extracted using Qiagen QIAquick gel extraction kit according to the manufacturer's instructions.

6.2.5 Restriction Digest

PCR product or plasmid to be digested (50 µL) was mixed with CutSmart buffer (6 µL), two appropriate restriction enzymes (2x 1 µL) and water (2 µL). The reaction was incubated at 37 °C for 3 h. PCR products were purified using Qiagen QIAquick PCR purification kit and the plasmid backbone was purified by gel electrophoresis and extraction with Qiagen QIAquick gel extraction kit.

6.2.6 Ligation

Plasmid backbone (50 ng) and insert containing the gene of interest (~75 ng, 1:3 molar ratio of backbone to insert) was mixed with T4 DNA ligase (1 µL) and 10x ligation buffer (1 µL) and made up to 10 µL with water. The reaction was incubated at room temperature for 15 min. 3 µL of the reaction was used to transform C2987 cells which were grown on LB agar + appropriate antibiotic.

6.2.7 Miniprep

Single colonies were picked from agar plates and inoculated in 10 mL LB + appropriate antibiotic. These cultures were shaken at 37 °C for ~18 h. In the morning, the cultures were harvested at 3500 xg in a Heraeus Function Line centrifuge for 10 min and the supernatant discarded. The plasmid DNA was extracted from the pellet using Qiagen QIAprep Miniprep kit according to the manufacturer's instructions.

6.2.8 Analytical Digest and Sequencing

Plasmid DNA from clones were checked for incorporation of the gene of interest by analytical digest. Plasmid DNA (7 µL) was mixed with CutSmart buffer (1 µL), 2x appropriate restriction enzyme (2x 1 µL) and incubated at 37 °C for 2 h. These digests were mixed with 6x purple DNA loading dye (NEB) and were analyzed by agarose gel electrophoresis. Positive clones have two bands, one with correct size for backbone and the other with correct size for gene of interest. Positive clones were sequenced by GATC using their T7 and pET-RP primers to determine if the gene of interest was correctly incorporated.

6.3 Protein Production

6.3.1 Transformation

Plasmid DNA (1 µL) was added to competent *E. coli* cells (25 µL) and incubated on ice for 20-30 min. The cells were subsequently heat shocked at 42 °C for 30 sec and replaced on ice for 2 min. SOC (100 µL) was added to the cells and these were recovered with shaking at 37 °C for 1 h. 40 µL of the recovered cells were grown on LB agar with appropriate antibiotic at 37 °C for ~18 h.

6.3.2 Gene Expression

Test expressions were carried out on a small scale to determine optimal conditions of expression prior to large scale expression. A single colony of *E. coli* cells containing an

expression plasmid was inoculated in 10 mL LB medium with appropriate antibiotic and grown for ~18 h at 37 °C with orbital shaking. The optical density of the cells at 600 nm (OD_{600}) was measured and this culture added to 250 mL LB with appropriate antibiotic for a final OD_{600} of 0.1. The new culture was grown to an OD_{600} between 0.6-0.8 at 37 °C with shaking. Sterile filtered stock solution of IPTG (0.5 M) was added to 18x 5 mL aliquots of the culture for a final concentration of 0.1 mM, 0.5 mM and 1 mM IPTG for six cultures each. Half of each of the cultures were further shaken at 16 and 30 °C respectively for 3 h, 5 h and ~18 h time points before harvesting by centrifugation at 3500 xg for 10 min. The cell pellet was resuspended in 1 mL of phosphate buffered saline (PBS) and lysed by sonication for 1.5 min (10 sec on and 10 sec off). The pellet and supernatant were separated by centrifugation at 24000 xg with a Heraeus Multifuge X3R equipped with an F15-8x50 rotor. Each fraction was tested by denaturing 12% or 15% SDS-PAGE gels run at 200 V. Optimal conditions for expression were determined and applied to large scale expression. The TamH protein was expressed in alternative LB media containing sorbitol to improve solubility of the protein (described in Section 6.1.3). Cells were harvested on a large scale at 4000 xg with a Heraeus Multifuge X3R equipped with a TX-1000 rotor for 30 min and cell pellets were stored at -20 °C prior to use.

6.4 Protein Purification

6.4.1 Cell Lysis

Cells were defrosted on ice, then resuspended in 30 mL Buffer A. The resulting solution was sonicated on ice for 15 min in 30 sec intervals in order to lyse the cells. The cell free extract was centrifuged at 24000 xg for 30 min and the supernatant filtered with a 0.45 µm filter. Protein was purified from the filtered supernatant.

6.4.2 Nickel NTA Purification

All purification steps were carried out at 4 °C or on ice. Proteins were purified by nickel immobilized metal affinity chromatography (IMAC) using a 1 mL HisTrap HP column attached to an ÄKTA Explorer which monitored the eluent at 280 nm. The supernatant from the cell

free extract was loaded onto the column at 1 mL/min, followed by washing of the column with Buffer A for 10 column volumes (CVs) to remove unbound protein. The protein of interest was eluted from the column with a gradient between Buffers A and B over 30 CVs. Eluted protein was concentrated to 5 mL and subjected to gel filtration chromatography (GFC) with the attached 6xHis tag in the case of proteins expressed in pET28a or pET22b. In the case of proteins expressed in pEHISTEV, the tag was removed by mixing the eluted protein with 1 mg of TEV protease and dialyzing this against Buffer D for 1 h. The resulting protein was purified by a second nickel IMAC step with the flow-through and wash fractions containing the cleaved protein. This protein was concentrated to 5 mL and subjected to GFC.

6.4.3 Gel Filtration Chromatography

HiLoad 16/60 Superdex S200 or Superdex S75 (for TamD ACP_{98} and TamA ACP) from GE healthcare were equilibrated with 1 CV of Buffer C on an ÄKTA Explorer. Proteins were eluted using 1 CV of Buffer C at 1 mL/min with 3 mL fractions collected and the eluent monitored at 280 nm. The fractions were analyzed by 12% or 15% denaturing SDS-PAGE gel. Those containing clean protein of interest were combined and concentrated for use or flash frozen at -80 °C.

6.4.4 Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis (SDS-PAGE)

A typical SDS-PAGE consisted of a 15% running gel: water (5.7 mL), 40% acrylamide (6 mL), 1.5 M Tris pH 8.8 (4 mL), 10 % (w/v) SDS (150 µL), 50 mg/mL ammonium persulfate (APS, 350 µL) and *N*,*N*,*N*',*N*'-tetramethylethylenediamine (TEMED, 20 µL) and a 4% stacking gel: water (2.9 mL), 40% acrylamide (0.75 mL), 0.5 M Tris pH 6.8 (1.25 mL), 10 % (w/v) SDS (50 µL), 50 mg/mL APS (100 µL) and TEMED (5 µL). Protein samples were denatured by the addition of 2x SDS loading buffer containing: 0.5 M Tris pH 6.8 (2.5 mL), glycerol (2 mL), 10% (w/v) SDS (4 mL), β-mercaptoethanol (1 mL) and 0.1 % (w/v) bromophenol blue (0.5 mL) and boiling for 10 min. Samples were loaded onto gels alongside protein molecular weight markers (PageRuler Unstained Protein Ladder, Thermo Fisher Scientific or Low Molecular Weight Marker, GE Healthcare) and were run for approximately 50 min at a constant voltage of 200

V. Subsequently gels were immersed in Expedeon Instant Blue Coomassie stain and incubated for 30 min at room temperature with gentle shaking.

6.5 Mass Spectrometry (MS)

6.5.1 Intact Protein LC ESI-MS

Bradford assay was used to determine protein concentration for liquid chromatography electrospray-MS (LC ESI-MS). Protein was diluted to a concentration of 10 µM and centrifuged for 10 min at 17000 xg. 5 µL of the supernatant was subjected to LC ESI-MS on a Synapt G2-Si Q-TOF (Waters) instrument with Phenomenex C4 3.6 µm LC column coupled to an ESI source. The LC gradient ran from 5% acetonitrile and water with 0.1% formic acid to 95% acetonitrile over 15 min. The MS source was set at 120 °C, backing pressure 2 mbar and sampling cone voltage of 54 V. Protein spectra are presented after subtraction using MassLynx V4.1 software. Peak *m/z* annotations were extracted from smoothed and centroided data which was also used in the component algorithm to extract the protein average masses.

6.5.2 4-Hydroxy-2,2'-Bipyrrole 5-Methanol (HBM) LC ESI-MS

Reactions producing HBM and its analogues were quenched with a 1:1 volume of acetonitrile with 0.01% trifluoroacetic acid (TFA) and centrifuged at 17000 xg. 5 µL of the supernatant was injected into a Phenomenex Kinetex 5 µm C18 100 Å column coupled to a micrOTOF II (Bruker). The LC gradient ran from 5% acetonitrile and 95% water with 0.1% formic acid to 95% acetonitrile over 10 min. Extracted ion chromatograms (EICs) and masses were determined on DataAnalysis V4.3 software.

6.5.3 Amines and Amides LC ESI-MS

Amine or amide producing reactions were quenched with a 1:1 volume of acetonitrile with 0.01% TFA and centrifuged at 17000 xg. 5 µL of supernatant was subjected to LC ESI-MS on a

Synapt G2-Si Q-TOF (Waters) instrument with Phenomenex Jupiter C18 5 µm 300 Å LC column coupled to an ESI source. The LC gradient ran from 5% acetonitrile and 95% water with 0.1% formic acid to 95% acetonitrile over 30 min. The MS source was set at 120 °C, backing pressure 2 mbar and sampling cone voltage of 54 V. Extracted ion chromatograms (EICs) and masses were determined on MassLynx V4.1 software.

6.5.4 Native MS

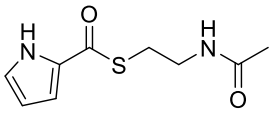
Protein was desalted into 300 mM ammonium bicarbonate using a P6 Micro BioSpin column to remove excess salt. Protein concentration was determined by Bradford assay and diluted to 10 µM. Protein was infused onto a Synapt G2-Si Q-TOF (Waters) instrument by nano-ESI using a TriVersa NanoMate (Advion) set at 1.7 kV and 0.8 psi. The MS source was set at 80 °C, backing pressure 5 mbar and sampling cone voltage of 200 V. The spectra were averaged over 4 min of acquisition and are presented without further data processing. Peak *m/z* annotations were extracted from smoothed and centroided data which was also used in the component algorithm to extract the protein average masses.

6.5.5 Peptide Digest and MS

50 µg of *holo*-TamA was denatured by addition of 8 M urea to a concentration of 6 M before incubation at room temperature for 2 h. Reduction of the protein was achieved with DTT added to a final concentration of 5 mM and further incubated at room temperature for 30 min. Iodoacetamide was added to a final concentration of 10 mM at room temperature in the dark for 30 min in order to acylate the protein. This reaction was quenched by adding additional DTT to a concentration of 10 mM with incubation at room temperature for 30 min. The sample was subsequently diluted to a urea concentration of less than 1 M with 50 mM ammonium bicarbonate. Finally, 1 µg of trypsin was added to the sample and the reaction was incubated at 37 °C for ~18 h. The samples were cleaned and concentrated using 100 µL C18 ziptips and washed in 0.2% formic acid. The peptides were eluted in 60% acetonitrile with 0.2% formic acid. The resulting peptide mixture was analyzed using a Bruker solariX X 12T in positive ion mode. The protein was infused using nano-ESI on a TriVersa NanoMate (Advion). Data were acquired over an *m/z* range of 200 – 5000 with an average of 30-50 scans

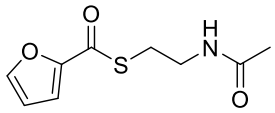
and an accumulation time of 0.2 sec. Resulting spectra were processed and calibrated with DataAnalysis V4.3. The $[M+3H]^{3+}$ ion containing the 4'-PP modification was isolated for further MS/MS analysis by increasing the accumulation time to 2 sec and averaging over 50 spectra which were acquired with a 2 Da window. The resulting product ion spectrum of $[M+3H]^{3+}$ was analyzed using ProsightLite[205] for y- and b- ions corresponding to a 4'-PP modification.

6.6 SNAC Synthesis



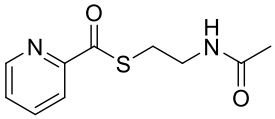
Pyrrole SNAC

S-(2-Acetamidoethyl) 1H-pyrrole-2-carbothioate: 2-Pyrrole carboxylic acid (500 mg, 4.5 mmol) and DMAP (137 mg, 1.1 mmol) were dissolved in dichloromethane (DCM, 50 mL). The mixture was stirred at 0 °C with the sequential addition of *N*-acetylcysteamine (0.44 mL, 4.5 mmol) and NEt_3 (0.69 mL, 5.0 mmol). EDC.HCl (1294 mg, 6.75 mmol) also dissolved in DCM was added dropwise. The mixture was stirred for ~18 h at room temperature under an Ar atmosphere. The reaction was quenched with saturated ammonium chloride and extracted with DCM. The organic phase was extracted, dried over $MgSO_4$, filtered and concentrated under reduced pressure. The crude product was purified by flash column chromatography with 95:5 DCM:MeOH. The colourless solid was obtained in 71% yield (674 mg). $\mathbf{R_f}$ (DCM:MeOH, 95:5) = 0.59; $\boldsymbol{\delta_H}$ (500 MHz, $CDCl_3$), 1.97 (3H, s, CH_3), 3.19 (2H, m, CH_2), 3.51 (2H, m, CH_2), 5.96 (1H, br s, NH), 6.29 (1H, m, CH), 7.03 (2H, m, CH), 9.19 (1H, s, NH), $\boldsymbol{\delta_C}$ (126 MHz, $CDCl_3$), 23.4 (CH_3), 27.8 (CH_2), 40.2 (CH_2), 111.2 (CH), 115.8 (CH), 124.1 (CH), 130.0 (C), 170.4 (CO), 181.8 (CO). ***mlz*** (ESI) for $[C_9H_{12}N_2O_2SNa]^+$ predicted: 235.0517, observed: 235.0540. Spectroscopic data is consistent with literature values.[178]



Furan SNAC

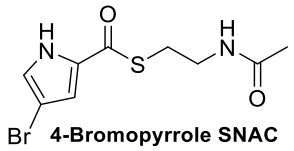
S-(2-Acetamidoethyl) furan-2-carbothioate: 2-Furoic acid (500 mg, 4.5 mmol) and DMAP (137 mg, 1.1 mmol) were dissolved in DCM (50 mL). The mixture was stirred at 0 °C with the sequential addition of *N*-acetylcysteamine (0.44 mL, 4.5 mmol) and NEt_3 (0.69 mL, 5.0 mmol). EDC.HCl (1294 mg, 6.75 mmol) also dissolved in DCM was added dropwise. The mixture was stirred for ~18 h at room temperature under an Ar atmosphere. The reaction was quenched with saturated ammonium chloride and extracted with DCM. The organic phase was extracted, dried over $MgSO_4$, filtered and concentrated under reduced pressure. The crude product was purified by flash column chromatography with 95:5 DCM:MeOH. The colourless solid was obtained in 80% yield (762 mg). $\mathbf{R_f}$ (DCM:MeOH, 95:5) = 0.52; $\boldsymbol{\delta_H}$ (500 MHz, $CDCl_3$), 1.97 (3H, s, CH_3), 3.21 (2H, t, *J* = 6.4 Hz, CH_2), 3.52 (2H, q, *J* = 6.0 Hz, CH_2), 5.96 (1H, br s, NH), 6.55 (1H, dd, *J* = 3.6, 1.7 Hz, CH), 7.21 (1H, dd, *J* = 3.6, 0.8 Hz, CH), 7.60 (1H, dd, *J* = 1.7, 0.8 Hz, CH), $\boldsymbol{\delta_C}$ (126 MHz, $CDCl_3$), 23.4 (CH_3), 28.0 (CH_2), 39.9 (CH_2), 112.5 (CH), 116.3 (CH), 146.7 (CH), 150.7 (C), 170.5 (CO), 180.8 (CO). ***mlz*** (ESI) for $[C_9H_{11}NO_3SNa]^+$ predicted: 236.0357, observed: 236.0358. Spectroscopic data is consistent with literature values.[178]



Pyridine SNAC

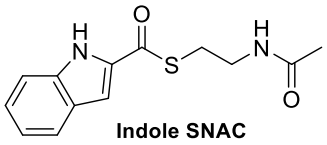
S-(2-Acetamidoethyl) pyridine-2-carbothioate: 2-Picolinic acid (554 mg, 4.5 mmol) and DMAP (137 mg, 1.1 mmol) were dissolved in DCM (50 mL). The mixture was stirred at room temperature with the sequential addition of *N*-acetylcysteamine (0.44 mL, 4.5 mmol) and NEt_3 (0.69 mL, 5.0 mmol). EDC.HCl (1294 mg, 6.75 mmol) also dissolved in DCM was added dropwise. The mixture was stirred for ~18 h at room temperature under an Ar atmosphere. The reaction was quenched with saturated ammonium chloride and extracted with DCM. The organic phase was extracted, dried over $MgSO_4$, filtered and concentrated under reduced pressure. The crude product was purified by flash column chromatography with 95:5 DCM:MeOH. The colourless solid was obtained in 74% yield (749 mg). $\mathbf{R_f}$ (DCM:MeOH, 95:5)

= 0.51; δ_H (500 MHz, $CDCl_3$), 1.97 (3H, s, CH_3), 3.21 (2H, t, *J* = 6.3 Hz, CH_2), 3.55 (2H, q, *J* = 6.3 Hz, CH_2), 5.88 (1H, br s, NH), 7.54 (1H, m, CH), 7.87 (1H, m, CH), 7.96 (1H, m, CH) 8.71 (1H, m, CH), δ_C (126 MHz, $CDCl_3$), 23.4 (CH_3), 28.7 (CH_2), 39.6 (CH_2), 120.7 (CH), 128.2 (CH), 137.5 (CH), 149.4 (CH), 151.8 (C), 170.3 (CO), 193.7 (CO). ***mlz*** (ESI) for $[C_{10}H_{12}N_2O_2SNa]^+$ predicted: 247.0517, observed: 247.0503. Spectroscopic data is consistent with literature values.[178]



4-Bromopyrrole SNAC

S-(2-Acetamidoethyl) 4-bromo-1H-pyrrole-2-carbothioate: Ethyl 4-bromopyrrole 2-carboxylate (300 mg, 1.38 mmol) was dissolved in methanol (10 mL) and KOH (20 M, 2 mL) and stirred at 50 °C for ~18 h. The solvent was removed under reduced pressure and HCl (0.5 M, 40 mL) was added to the residue. This solution was partitioned in diethyl ether and the organic fraction dried over $MgSO_4$. The solvent was again removed under reduced pressure to yield 4-bromopyrrole 2-carboxylic acid (244.5 mg, 94% yield) which was used in the subsequent step without further purification. 4-Bromo 2-pyrrole carboxylic acid (239 mg, 1.25 mmol) and DMAP (38 mg, 0.3 mmol) were dissolved in DCM (30 mL). The mixture was stirred at room temperature with the sequential addition of *N*-acetylcysteamine (0.12 mL, 1.25 mmol) and NEt_3 (0.19 mL, 1.4 mmol). EDC.HCl (360 mg, 1.88 mmol) also dissolved in DCM was added dropwise. The mixture was stirred for ~18 h at room temperature under an Ar atmosphere. The reaction was quenched with saturated ammonium chloride and extracted with DCM. The organic phase was extracted, dried over $MgSO_4$, filtered and concentrated under reduced pressure. The crude product was purified by flash column chromatography with 95:5 DCM:MeOH. The cream coloured solid was obtained in 47% yield (173 mg). R_f (DCM:MeOH, 95:5) = 0.30; δ_H (500 MHz, $CDCl_3$), 1.97 (3H, s, CH_3), 3.20 (2H, t, *J* = 6.3 Hz, CH_2), 3.51 (2H, q, *J* = 5.9 Hz, CH_2), 5.89 (1H, s, NH), 7.01 (2H, m, CH), 9.23 (1H, s, NH), δ_C (126 MHz, $CDCl_3$), 23.3 (CH_3), 28.0 (CH_2), 39.8 (CH_2), 98.5 (CBr), 116.9 (CH), 123.6 (CH), 130.0 (C), 170.3 (CO), 181.2 (CO). ***mlz*** (ESI) for $[C_9H_{12}BrN_2O_2S]^+$ predicted: 290.9803 (97%), 292.9702 (100%), observed: 290.9772 (91%), 292.9759 (100%).



Indole SNAC

S-(2-Acetamidoethyl) 1H-indole-2-carbothioate: 2-Indole carboxylic acid (725 mg, 4.5 mmol) and DMAP (137 mg, 1.1 mmol) were dissolved in DCM (50 mL). The mixture was stirred at room temperature with the sequential addition of *N*-acetylcysteamine (0.44 mL, 4.5 mmol) and NEt_3 (0.69 mL, 5.0 mmol). EDC.HCl (1294 mg, 6.75 mmol) also dissolved in DCM was added dropwise. The mixture was stirred for ~18 h at room temperature under an Ar atmosphere. The reaction was quenched with saturated ammonium chloride and extracted with DCM. The organic phase was extracted, dried over $MgSO_4$, filtered and concentrated under reduced pressure. The crude product was purified by flash column chromatography with 95:5 DCM:MeOH. The colourless crystalline solid was obtained in 19% yield (230 mg). $\mathbf{R_f}$ (DCM:MeOH, 95:5) = 0.35; $\boldsymbol{\delta_H}$ (500 MHz, $CDCl_3$), 1.99 (3H, s, CH_3), 3.27 (2H, m, CH_2), 3.55 (2H, m, CH_2), 5.91 (1H, s, NH), 7.18 (1H, m, CH), 7.36 (2H, m, CH), 7.42 (1H, m, CH), 7.71 (1H, m, CH), 8.90 (1H, s, CH), $\boldsymbol{\delta_C}$ (126 MHz, $CDCl_3$), 23.3 (CH_3), 28.2 (CH_2), 39.9 (CH_2), 108.9 (CH), 112.0 (CH), 121.3 (CH), 123.0 (CH), 126.5 (CH), 127.3 (C), 133.6 (C), 136.8 (C), 170.3 (CO), 184.0 (CO). ***mlz*** (ESI) for $[C_{13}H_{14}N_2O_2SNa]^+$ predicted: 285.0674, observed: 285.0633.

6.7 Enzymatic Reactions

6.7.1 ACP Post-Translational Modification

Where indicated TamD, TamD ACP, TamA and TamA ACP were 4'-phosphopantetheine (4'-PP) modified during the purification process. After the TEV protease cleavage and dialysis, the protein was concentrated and added to a 10 mL reaction containing 10 µM Sfp, 10 mM MgCl2 and 500 µM CoA in Buffer D. This reaction was left shaking for ~18 h at 4 °C. The protein purification was subsequently continued in the same way with the second nickel affinity chromatography step removing the 6xHis tagged Sfp along with the TEV protease. Alternatively, the proteins were 4'-PP modified *in vivo* by coexpression with untagged Sfp in pACYC.

6.7.2 HBM Reactions

Reactions producing HBM or any of its intermediates and analogues were carried out in Buffer C with 50 µM PLP and reactants in the following concentrations: 10 µM TamF, 2 mM SNAC, 500 µM malonyl-CoA, 5 µM *E. coli* FabD and 50 mM amino acid. Reaction time for generation of intermediates was 15 min at room temperature. Detection of the intermediates was carried out by intact protein MS as they are covalently protein bound. HBM and its analogues were produced in reactions incubated at room temperature that were quenched with a 1:1 volume of 0.01% TFA in acetonitrile at various time points (0 min, 2 min, 5 min, 10 min, 15 min, 30 min, 1 h, 2 h, 3 h and 5 h). These samples were analyzed by LC ESI-MS and high performance liquid chromatography (HPLC).

6.7.3 TamA and TamA ACP Transthioesterification

Holo-TamA ACP was mixed 1:1 with as purified *apo*-TamA at a final concentration of 10 µM of each protein in Buffer C. The mixture was incubated for ~18 h at room temperature before being subjected to LC ESI-MS to determine the fatty acids bound in the TamA active site after purification. To test the transfer of alternative fatty acids, 10 µM *holo*-TamA and 10 µM *holo*-TamA ACP were incubated with 1 mM fatty acid (from a 10 mM stock in DMSO), 5 mM ATP and 10 mM $MgCl_2$ in buffer C for 20 h at room temperature. The reactions were subsequently subjected to LC ESI-MS.

6.7.4 TamH Reduction

In order to determine whether the thioester reductase (TR) domain of TamH utilizes NADH or NADPH as its cofactor, reactions with each substrate were monitored at 340 nm. The reactions were carried out in Buffer C and contained 10 µM *holo*-TamA, 1.25 mM C12 lauric acid (from a 25 mM stock in DMSO), 1.25 mM ATP, 10 mM $MgCl_2$, 10 µM TamH, 250 µM PLP and 160 µM NAD(P)H. Control reactions were carried out without enzyme. These were monitored continuously on a Varian Cary® UV/Vis spectrophotometer at 340nm for 10 min.

6.7.5 TamH Transamination

The TamH transamination reaction contained 10 µM TamH, 250 µM PLP, 5 mM L-glutamic acid and 1 mM C12 aldehyde (from a 10 mM stock in DMSO) in Buffer C. The reactions were incubated at 37 °C for 24 h before being analyzed by LC ESI-MS.

6.7.6 Amine Reactions

The reactions producing dodecamine from C12 lauric acid by TamA and TamH were carried out in Buffer C and contained 10 µM *holo*-TamA, 1 mM C12 lauric acid (from a 25 mM stock in DMSO), 5 mM ATP, 10 mM $MgCl_2$, 10 µM TamH, 250 µM PLP and 2 mM NAD(P)H. Reactions substituting the TamA protein with C12-CoA contained 1 mM C12-CoA, 10 µM TamH, 250 µM PLP and 2 mM NAD(P)H. Reactions were carried out with both NADH and NADPH separately along with a negative control without the TamH enzyme. The reactions were incubated at 37 °C for ~18 h before the addition of 5 mM L-glutamic acid to allow transamination. The reactions were then further incubated for 3 h at 37 °C.

6.7.7 ANL Reactions

Reactions were carried out according to the methodology of the Flitsch group.[210] They contained 10 µg ANL, 1 mM fatty acid (from a 25 mM stock in DMSO), 5 mM ATP, 10 mM $MgCl_2$ and 100 mM amine (pH 9) in 100 mM sodium carbonate-sodium bicarbonate buffer (pH 9).

6.8 UV/Visible Spectroscopy

6.8.1 PLP Enzyme-Amino Acid Dissociation Constants (K_d)

UV/visible spectroscopy was carried out on a Varian Cary® UV/Vis spectrophotometer. Excess PLP and Tris base was removed by PD10 column (GE healthcare) and the enzyme exchanged into Buffer E. Assays were carried out in 1 cm pathlength cuvettes and baseline correction was carried out before acquiring spectra. Protein was diluted to a concentration ~40 µM and

amino acid substrates were added from a 1 M stock solution to various final concentrations (0-100 mM). Enzyme and substrate mixtures were allowed to equilibrate at room temperature for 5 min before spectra were acquired. The spectra were normalized against the 280 nm peak to account for dilution of the sample with addition of substrate solution. Changes in the absorbance maximum of the ketoenamine peak (~420 nm) were plotted and fitted with a hyperbolic saturation curve using Origin software. The dissociation constant (K_d) was calculated according to the following equation:

$$\Delta A_{obs} = \frac{\Delta A_{max}\,[\text{amino acid}]}{K_d + [\text{amino acid}]}$$

Where ΔA_{obs} is the observed change in absorbance at ~420 nm, ΔA_{max} is the maximal absorbance change and [amino acid] is the concentration of the added amino acid substrate.

6.8.2 High Performance Liquid Chromatography (HPLC)

HBM producing reactions were quenched by addition of a 1:1 volume of acetonitrile with 0.01% TFA. The samples were centrifuged at 17000 xg for 10 min before injection of 10 µL of the supernatant onto a Phenomenex Luna 5 µm C18 100 Å HPLC column. Samples were eluted with water and 0.1% TFA for 5 min followed by a 15 min gradient to 55% acetonitrile and 0.1% TFA. This concentration was maintained for 5 min and then returned to 100% water and 0.1% TFA for an additional 5 min. The eluent was monitored at 345 nm.

6.9 X-ray Crystallography

Structural studies were carried out in the Scottish Structural Proteomics Facility (SSPF) in collaboration with Dr. Stephen McMahon in the Naismith group at the University of St. Andrews. Proteins were concentrated by Vivaspin column to (10-21 mg/mL) and centrifuged at 17000 xg for 10 min to remove particulate. Crystallization conditions were screened by a nano-drop crystallization robot (Gryphon). This robot sets up sitting drop 96-well plate crystal trials with three different ratios of protein: precipitant 1:1, 2:1 and 1:2 for drop sizes of 0.2 or 0.3 µL. The plates were left at 20 °C for several weeks before checking for the formation of crystals.

7 Conclusions and Future Work

Tambjamine YP1 is an important bipyrrole natural product with an attached fatty amine tail that has antibacterial and cytotoxic activity. Understanding of the biosynthetic pathway to this molecule is important not only for the production of improved tambjamines but also for the continued exploration of natural product biosynthetic enzymes. This work encompasses the first detailed research into the activity and mechanism of the enzymes involved in tambjamine YP1 biosynthesis in *P. tunicata*. A summary of the enzymes studied and their roles in the biosynthetic pathway are shown in Figure 7.1.

Recombinant forms of TamF and TamD have been cloned, expressed, purified and their roles in the biosynthesis of the bipyrrole moiety of tambjamine YP1 have been determined. MS analysis of the interesting CLF-KS didomain protein TamF illustrated that covalent attachment of a pyrrole substrate occurs as the first step of the catalyzed reaction. The native PCP-pyrrole substrate can be substituted with a pyrrole SNAC mimic for simple loading of this protein. TamD, the ACP-AOS didomain enzyme must be loaded with its 4'-PP arm for activity, which can be achieved *in vitro* or *in vivo* with *B. subtilis* Sfp. Further loading of the protein with a malonyl group was carried out in a novel manner with *E. coli* FabD to allow for multiple turnovers of this enzyme. MS analysis of the stand-alone TamD ACP_{98} domain has shown that the TamF KS can subsequently catalyze the decarboxylation of malonyl-TamD ACP and attack of its thioester bound substrate to produce the diketopyrrole-ACP thioester. This thioester is turned over in the presence of L-serine by the catalytic TamD AOS domain to produce a linear α-oxoamine product. Spontaneous cyclisation then occurs to form the first bipyrrole intermediate, HBM, which has been detected for the first time *in vitro* using a combination of UV/visible spectroscopy, HPLC and MS analysis.

This work represents an advance in the understanding of bipyrrole biosynthesis not just for the tambjamine family but also for the prodigiosins and other polyketides. In previous work on the MBC based natural products HBM had never been identified as a product of the MBC biosynthetic enzymes. This was likely due to the limitation in the previously discussed assay method which only allowed one turnover of the reaction system. Therefore, this is the first observation of HBM as an intermediate in the production of this family of natural products. It has also been the first exploration of the substrate scope of the bipyrrole forming enzymes.

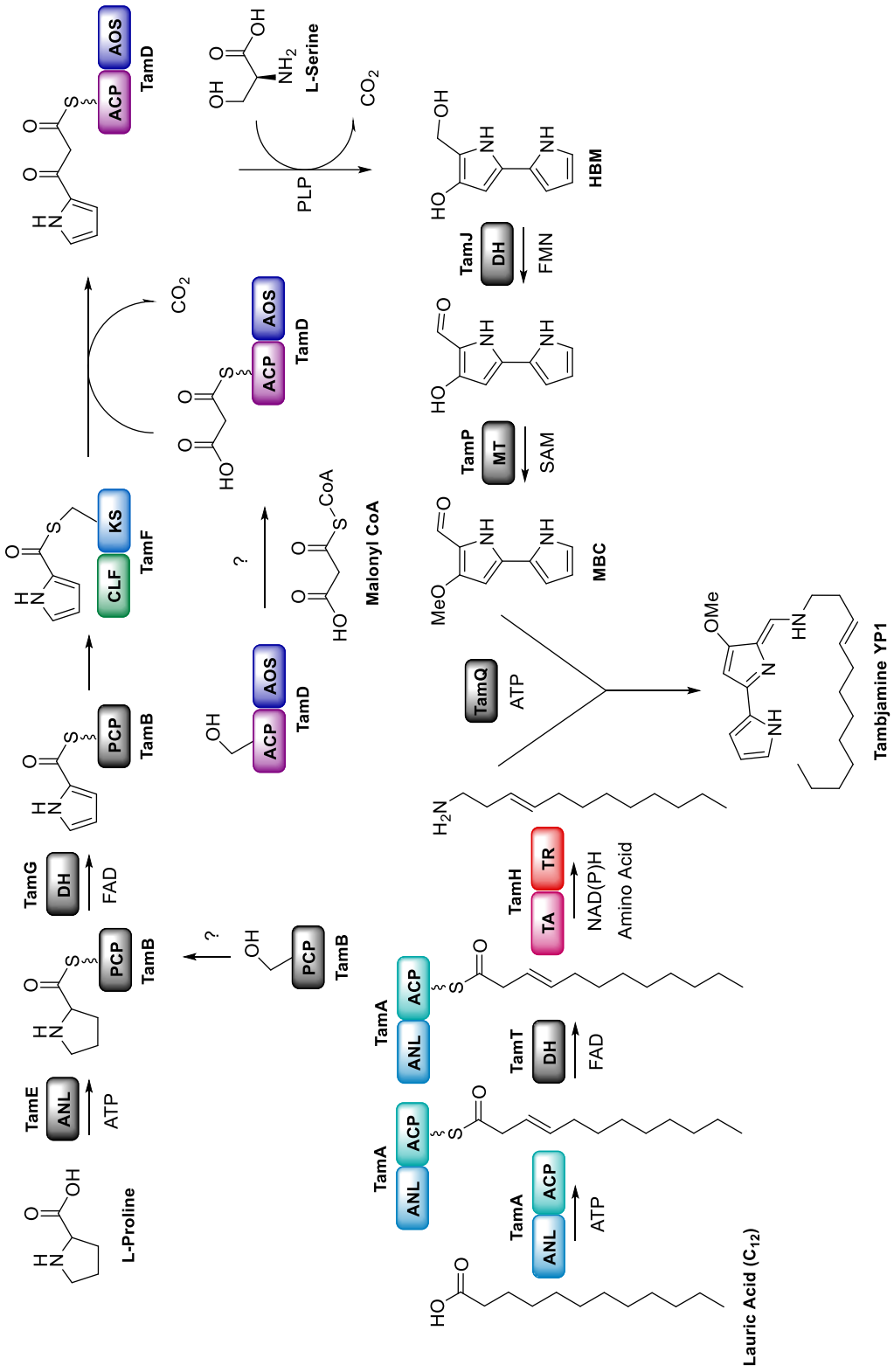


Figure 7.1: Updated biosynthetic pathway to tambjamine YP1 from L-proline, malonyl-CoA, L-serine and lauric acid in *P. tunicata,* with the enzymes and transformations studied in this work indicated by the coloured domains, corrected from Burke *et al.*[50]

The potential for these proteins to accept a range of substrates may be low; of the few amino acids tested L-alanine was the only other amino acid accepted in the reaction and none of the alternative heterocycles could be utilized. However, only a limited number of substrates were employed in the system and a more comprehensive review of potential substrates is needed. Determination of 3D structures of these enzymes could also lead to engineering of the system to accept novel substrates. Therefore, this elucidated pathway opens the door to the production of novel tambjamines as well as other unusual bipyrrole or bicyclic systems.

The previously unexplored pathway to the tambjamine YP1 amine tail has also been elucidated for the first time and the predicted biosynthetic route has been revised. A gene in the tambjamine YP1 cluster without a previously hypothesized function (*tamA*) has proved to be the protein necessary for activation of C12 fatty acid. This protein is an ANL-ACP fusion which creates a fatty acid adenylate using ATP, activating the carboxylic acid for attack by the 4'-PP modification of its C-terminal ACP. Transthioesterification reactions between the full-length TamA and a truncated TamA ACP domain have shown the substrate specificity of the protein for saturated fatty acids to range from C6-C13 in chain length. TamA was also coupled with the activity of the downstream TA-TR didomain enzyme TamH to show that together these two proteins can convert a fatty acid to the corresponding alkyl amine. The TamA ACP-bound fatty acid thioester is reduced by TamH using NADH to produce an aldehyde intermediate. The aldehyde is subsequently transaminated using L-glutamic acid to yield the final amine product. In contrast to the previous hypothesis, this work has proven that TamA is the activating enzyme required for the biosynthesis of the tambjamine YP1 amine tail. TamH, a unique enzyme fusion has also been explored for the first time. However, full characterization of this pathway will also require the soluble expression and purification of TamT, the DH domain that generates the 3,4 alkene moiety. Future work will require optimization of this reaction system as well as determination of the kinetic parameters. This will allow its exploration as a potential biocatalytic cascade for alkyl amine production as well as the production of novel tambjamine derivatives with alternative hydrophobic tails.

Finally, the TamA ANL domain was truncated from its ACP partner in an attempt to use this protein for a biocatalytic application. The adenylation domain is able to create fatty acid adenylate intermediates which can be attacked by an amine molecule to yield an amide. Simple amides were produced from C9-C13 fatty acids with ammonia. As the C12 fatty acid was the best fatty acid substrate for this reaction, it was used to test the amine specificity of

the reaction. The TamA ANL domain could accept all the amines tested including L-glutamic acid, D-glutamic acid, L-aspartic acid, L-histidine, ethanolamine and benzylamine. Therefore, this reaction does not have enantioselectivity, suggesting that the ANL domain is not catalyzing the amide bonding forming reaction. Instead, the amine enters the active site without binding and traps the already activated intermediate. However, this data shows that this enzyme has broad substrate scope and is an excellent addition to the biocatalytic tool box. Future work on this enzyme should include scale up and determination of kinetic parameters for the production of the interesting *N*-acylhistidine molecules produced by the human pathogen *L. pneumophila.* Determination of the TamA ANL structure would also allow the engineering of this enzyme to accept longer chains for the synthesis of other high value products such as PEA or fluorescent fatty acids.

This thesis has explored the convergent biosynthesis of the two building blocks of the important, biologically active tambjamine YP1 molecule. The pathway to each fragment has been elucidated and has shed light on the methods of bipyrrole and alkyl amine biosynthesis. Natural product enzymes have often been used for biocatalytic application due to their remarkable and specific chemistry. One of the Tam enzymes has already been exploited for this purpose and the characterization of these proteins has opened the door to further development of this pathway.

8 References

(1) Newman, D. J.; Cragg, G. M. Natural Products as Sources of New Drugs from 1981 to 2014. *J. Nat. Prod.* **2016**, *79* (3), 629–661.

(2) Weissman, K. J.; Leadlay, P. F. Combinatorial Biosynthesis of Reduced Polyketides. *Nat. Rev.* **2005**, *3*, 925–936.

(3) Wender, P. A.; Miller, B. L. Synthesis at the Molecular Frontier. *Nat. Horizons* **2009**, *460*, 197–201.

(4) Morrison, K. C.; Hergenrother, P. J. Natural Products as Starting Points for the Synthesis of Complex and Diverse Compounds. *Nat. Prod. Rep.* **2014**, *31*, 6–14.

(5) Walsh, C. T.; Fischbach, M. A. Natural Products Version 2.0: Connecting Genes to Molecules. *J. Am. Chem. Soc.* **2010**, *132*, 2469–2493.

(6) Li, J. W.-H.; Vederas, J. C. Drug Discovery and Natural Products: End of an Era or an Endless Frontier? *Science* **2009**, *325* (5937), 161–165.

(7) Traxler, M. F.; Kolter, R. Natural Products in Soil Microbe Interactions and Evolution. *Nat. Prod. Rep.* **2015**, *32*, 956–970.

(8) Nichols, D.; Cahoon, N.; Trakhtenberg, E. M.; Pham, L.; Mehta, A.; Belanger, A.; Kanigan, T.; Lewis, K.; Epstein, S. S. Use of Ichip for High-Throughput in Situ Cultivation of "Uncultivable" Microbial Species. *Appl. Environ. Microbiol.* **2010**, *76* (8), 2445–2450.

(9) Stonik, V. A. Marine Natural Products : A Way to New Drugs. *Acta Naturae* **2009**, *2*, 15–25.

(10) Challis, G. L. Genome Mining for Novel Natural Product Discovery. *J. Med. Chem.* **2008**, *51*, 2618–2628.

(11) Harvey, A. L.; Edrada-Ebel, R.; Quinn, R. J. The Re-Emergence of Natural Products for Drug Discovery in the Genomics Era. *Nat. Rev.* **2015**, *14* (2), 111–129.

(12) Walsh, C. T. Nature Loves Nitrogen Heterocycles. *Tetrahedron Lett.* **2015**, *56*, 3075–3081.

(13) Walsh, C. T.; Garneau-Tsodikova, S.; Howard-Jones, A. R. Biological Formation of Pyrroles: Nature's Logic and Enzymatic Machinery. *Nat. Prod. Rep.* **2006**, *23*, 517–531.

(14) Poulos, T. L. Heme Enzyme Structure and Function. *Chem. Rev.* **2014**, *114*, 3919–3962.

(15) Hooper, D. C.; Wolfson, J. S.; McHugh, G. L.; Winters, M. B.; Swartz, M. N. Effects of Novobiocin, Coumermycin A1, Clorobiocin, and Their Analogs on Escherichia Coli DNA Gyrase and Bacterial Growth. *Antimicrob. Agents Chemother.* **1982**, *22* (4), 662–671.

(16) Doi, K.; Gowda, K.; Liu, Q.; Lin, J.-M.; Sung, S.-S.; Dower, C.; Claxton, D.; Loughran Jr, T. P.; Amin, S.; Wang, H.-G. Pyoluteorin Derivatives Induce Mcl-1 Degradation and Apoptosis in Hematological Cancer Cells. *Cancer Biol. Ther.* **2014**, *15* (12), 1688–1699.

(17) Sáez Díaz, R. I.; Regourd, J.; Santacroce, P. V.; Davis, J. T.; Jakeman, D. L.; Thompson, A. Chloride Anion Transport and Copper-Mediated DNA Cleavage by C-Ring Functionalized Prodigiosenes. *Chem. Commun.* **2007**, *26*, 2701–2703.

(18) Shinkre, B. A.; Raisch, K. P.; Fan, L.; Velu, S. E. Analogs of the Marine Alkaloid Makaluvamines: Synthesis, Topoisomerase II Inhibition, and Anticancer Activity. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 2890–2893.

(19) Raju, R.; Piggott, A. M.; Diaz, L. X. B.; Khalil, Z.; Capon, R. J. Heronapyrroles A-C: Farnesylated 2-Nitropyrroles from an Australian Marine-Derived Streptomyces Sp. *Org. Lett.* **2010**, *12* (22), 5158–5161.

(20) Galm, U.; Heller, S.; Shapiro, S.; Page, M.; Li, S.-M.; Heide, L. Antimicrobial and DNA Gyrase-Inhibitory Activities of Novel Clorobiocin Derivatives Produced by Mutasynthesis. *Antimicrob. Agents Chemother.* **2004**, *48* (4), 1307–1312.

(21) Chawrai, S. R.; Williamson, N. R.; Salmond, G. P. C.; Leeper, F. J. Chemoenzymatic Synthesis of Prodigiosin Analogues - Exploring the Substrate Specificity of PigC. *Chem. Commun.* **2008**, *0*, 1862–1864.

(22) Roth, B. D. The Discovery and Development of Atorvastatin, a Potent Novel Hypolipidemic Agent. *Prog. Med. Chem.* **2002**, *40*, 1–22.

(23) Cordrey, L. J. Tolmetin Sodium, a New Anti-Arthritis Drug: Double-Blind and Long-Term Studies. *J. Am. Geriatr. Soc.* **1976**, *24* (10), 440–446.

(24) Demetri, G. D.; van Oosterom, A. T.; Garrett, C. R.; Blackstein, M. E.; Shah, M. H.; Verweij, J.; McArthur, G.; Judson, I. R.; Heinrich, M. C.; Morgan, J. A.; et al. Efficacy and Safety of Sunitinib in Patients with Advanced Gastrointestinal Stromal Tumour after Failure of Imatinib: A Randomised Controlled Trial. *Lancet* **2006**, *368*, 1329–1338.

(25) Desai, A. S. Single-Dose Treatment of Oxyuriasis with Pyrvinium Embonate. *Br. Med. J.* **1962**,

1583–1585.

(26) Downey, A. S.; Chong, C. R.; Graczyk, T. K.; Sullivan, D. J. Efficacy of Pyrvinium Pamoate against Cryptosporidium Parvum Infection In Vitro and in a Neonatal Mouse Model. *Antimicrob. Agents Chemother.* **2008**, *52* (9), 3106–3112.

(27) Hu, D. X.; Withall, D. M.; Challis, G. L.; Thomson, R. J. Structure, Chemical Synthesis, and Biosynthesis of Prodiginine Natural Products. *Chem. Rev.* **2016**, *116*, 7818–7849.

(28) Carté, B.; Faulkner, D. J. Defensive Metabolites from Three Nembrothid Nudibranchs. *J. Org. Chem.* **1982**, *48* (14), 2314–2318.

(29) Carté, B.; Faulkner, D. J. Role of Secondary Metabolites in Feeding Associations between a Predatory Nudibranch, Two Grazing Nudibranchs, and a Bryozoan. *J. Chem. Ecol.* **1986**, *12* (3), 795–804.

(30) Linquist, N.; Fenical, W. New Tambjamine Class Alkaloids from the Marine Ascidian Atapozoa Sp. and Its Nudibranch Predators. Origin of the Tambjamines in Atapozoa. *Experientia* **1991**, *47*, 504–506.

(31) Paul, V. J.; Lindquist, N.; Fenical, W. Chemical Defenses of the Tropical Ascidian Atapozoa Sp. and Its Nudibranch Predators Nembrotha Spp. *Mar. Ecol. Prog. Ser.* **1990**, *59*, 109–118.

(32) Blackman, A. J.; Li, C. New Tambjamine Alkaloids from the Marine Bryozoan Bugula Dentata. *Aust. J. Chem.* **1994**, *47*, 1625–1629.

(33) Kojiri, K.; Nakajima, S.; Suzuki, H.; Okura, A.; Suda, H. A New Antitumor Substance, BE-18591, Produced by a Streptomycete I. *J. Antibiot. (Tokyo).* **1993**, *46* (12), 1799–1803.

(34) Nakajima, S.; Kojiri, K.; Suda, H. A New Antitumor Substance, BE-18591, Produced by a Streptomycete II. *J. Antibiot. (Tokyo).* **1993**, *46* (12), 1894–1896.

(35) Franks, A.; Haywood, P.; Holstrom, C.; Egan, S.; Kjelleberg, S.; Kumar, N. Isolation and Structure Elucidation of a Novel Yellow Pigment from the Marine Bacterium Pseudoalteromonas Tunicata. *Molecules* **2005**, *10*, 1286–1291.

(36) Pinkerton, D. M.; Banwell, M. G.; Garson, M. J.; Kumar, N.; Odorico de Moraes, M.; Cavalcanti, B. C.; Barros, F. W. A.; Pessoa, C. Antimicrobial and Cytotoxic Activities of Synthetically Derived Tambjamines C and E – J, BE-18591, and a Related Alkaloid from the Marine Bacterium Pseudoalteromonas Tunicata. *Chem. Biodivers.* **2010**, *7*, 1311–1324.

(37) Melvin, M. S.; Ferguson, D. C.; Lindquist, N.; Manderville, R. A. DNA Binding by 4-Methoxypyrrolic Natural Products. Preference for Intercalation at AT Sites by Tambjamine E and Prodigiosin. *J. Org. Chem.* **1999**, *64*, 6861–6869.

(38) Melvin, M. S.; Wooton, K. E.; Rich, C. C.; Saluta, G. R.; Kucera, G. L.; Lindquist, N.; Manderville, R. A. Copper-Nuclease Efficiency Correlates with Cytotoxicity for the 4-Methoxypyrrolic Natural Products. *J. Inorg. Biochem.* **2001**, *87*, 129–135.

(39) Cavalcanti, B. C.; Júnior, H. V. N.; Seleghim, M. H. R.; Berlinck, R. G. S.; Cunha, G. M. a; Moraes, M. O.; Pessoa, C. Cytotoxic and Genotoxic Effects of Tambjamine D, an Alkaloid Isolated from the Nudibranch Tambja Eliora, on Chinese Hamster Lung Fibroblasts. *Chem. Biol. Interact.* **2008**, *174* (3), 155–162.

(40) Aldrich, L. N.; Stoops, S. L.; Crews, B. C.; Marnett, L. J.; Lindsley, C. W. Total Synthesis and Biological Evaluation of Tambjamine K and a Library of Unnatural Analogs. *Bioorg. Med. Chem. Lett.* **2010**, *20*, 5207–5211.

(41) Kotev, M.; Manuel-Manresa, P.; Hernando, E.; Soto-Cerrato, V.; Orozco, M.; Quesada, R.; Perez-Tomas, R.; Guallar, V. Inhibition of Human Enhancer of Zeste Homolog 2 (EZH2) with Tambjamine Analogs. *J. Chem. Inf. Model.* **2017**, *58* (8), 2089–2098.

(42) Iglesias Hernández, P.; Moreno, D.; Javier, A. A.; Torroba, T.; Pérez-Tomás, R.; Quesada, R. Tambjamine Alkaloids and Related Synthetic Analogs: Efficient Transmembrane Anion Transporters. *Chem. Commun.* **2012**, *48* (10), 1556–1558.

(43) Saggiomo, V.; Otto, S.; Marques, I.; Félix, V.; Torroba, T.; Quesada, R. The Role of Lipophilicity in Transmembrane Anion Transport. *Chem. Commun.* **2012**, *48* (43), 5274–5276.

(44) Hernando, E.; Soto-Cerrato, V.; Cortés-Arroyo, S.; Pérez-Tomás, R.; Quesada, R. Transmembrane Anion Transport and Cytotoxicity of Synthetic Tambjamine Analogs. *Org. Biomol. Chem.* **2014**, *12* (11), 1771–1778.

(45) Kancharla, P.; Kelly, J. X.; Reynolds, K. A. Synthesis and Structure–Activity Relationships of Tambjamines and B-Ring Functionalized Prodiginines as Potent Antimalarials. *J. Med. Chem.* **2015**, *58*, 7286–7309.

(46) Holmstrom, C.; James, S.; Neilan, B. A.; White, D. C.; Kjelleberg, S. Pseudoalteromonas Tunicata Sp. Nov., a Bacterium That Produces Antifouling Agents. *Int. J. Syst. Bacteriol.* **1998**, *48*, 1205–1212.

(47) Thomas, T.; Evans, F. F.; Schleheck, D.; Mai-Prochnow, A.; Burke, C.; Penesyan, A.; Dalisay, D. S.; Stelzer-Braid, S.; Saunders, N.; Johnson, J.; et al. Analysis of the Pseudoalteromonas Tunicata Genome Reveals Properties of a Surface-Associated Life Style in the Marine Environment. *PLoS One* **2008**, *3* (9), e3252.

(48) Holmstrom, C.; Egan, S.; Franks, A.; McCloy, S.; Kjelleberg, S. Antifouling Activities Expressed by Marine Surface Associated Pseudoalteromonas Species. *FEMS Microbiol. Ecol.* **2002**, *41*, 47–58.

(49) Lichstein, H. C.; van de Sand, V. F. Violacein, an Antibiotic Pigment Produced by Chromobacterium Violaceum. *J. Infect. Dis.* **1945**, *76* (1), 46–51.

(50) Burke, C.; Thomas, T.; Egan, S.; Kjelleberg, S. The Use of Functional Genomics for the Identification of a Gene Cluster Encoding for the Biosynthesis of an Antifungal Tambjamine in the Marine Bacterium Pseudoalteromonas Tunicata. *Environ. Microbiol.* **2007**, *9* (3), 814–818.

(51) Altschup, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. Basic Local Alignment Search Tool. *J. Mol. Biol.* **1990**, *215*, 403–410.

(52) Williamson, N. R.; Fineran, P. C.; Leeper, F. J.; Salmond, G. P. C. The Biosynthesis and Regulation of Bacterial Prodiginines. *Nat. Rev.* **2006**, *4*, 887–899.

(53) Crosby, J.; Crump, M. P. The Structural Role of the Carrier Protein - Active Controller or Passive Carrier. *Nat. Prod. Rep.* **2012**, *29*, 1111–1137.

(54) Stanley, A. E.; Walton, L. J.; Kourdi Zerikly, M.; Corre, C.; Challis, G. L. Elucidation of the Streptomyces Coelicolor Pathway to 4-Methoxy-2,2'-bipyrrole-5-Carboxaldehyde, an Intermediate in Prodiginine Biosynthesis. *Chem. Commun.* **2006**, No. 38, 3981–3983.

(55) Garneau-Tsodikova, S.; Dorrestein, P. C.; Kelleher, N. L.; Walsh, C. T. Protein Assembly Line Components in Prodigiosin Biosynthesis: Characterisation of PigA,G,H,I,J. *J. Am. Chem. Soc.* **2006**, *128*, 12600–12601.

(56) Haynes, S. W.; Sydor, P. K.; Stanley, A. E.; Song, L.; Challis, G. L. Role and Substrate Specificity of the Streptomyces Coelicolor RedH Enzyme in Undecylprodiginine Biosynthesis. *Chem. Commun.* **2008**, *0*, 1865–1867.

(57) Chawrai, S. R.; Williamson, N. R.; Mahendiran, T.; Salmond, G. P. C.; Leeper, F. J. Characterisation of PigC and HapC, the Prodigiosin Synthetases from Serratia Sp. and Hahella Chejuensis with Potential for Biocatalytic Production of Anticancer Agents. *Chem. Sci.* **2012**, *3*,

447.

(58) Cortes, J.; Haydock, S. F.; Roberts, G. A.; Bevitt, D. J.; Leadlay, P. F. An Unusually Large Multifunctional Polypeptide in the Erythromycin-Producing Polyketide Synthase of Saccharopolyspora Erythraea. *Nature* **1990**, *348*, 176–178.

(59) Donadio, S.; Staver, M. J.; McAlpine, J. B.; Swanson, S. J.; Katz, L. Molecular Organisation of Genes Required for Complex Polyketide Biosynthesis. *Science* **1991**, *252*, 675–679.

(60) Mootz, H. D.; Marahiel, M. A. The Tyrocidine Biosynthesis Operon of Bacillus Brevis: Complete Nucleotide Sequence and Biochemical Characterization of Functional Internal Adenylation Domains. *J. Bacteriol.* **1997**, *179* (21), 6843–6850.

(61) Weissman, K. J. The Structural Biology of Biosynthetic Megaenzymes. *Nat. Chem. Biol.* **2015**, *11*, 660–670.

(62) Marahiel, M. A.; Stachelhaus, T.; Mootz, H. D. Modular Peptide Synthetases Involved in Nonribosomal Peptide Synthesis. *Chem. Rev.* **1997**, *97*, 2651–2673.

(63) Chan, D. I.; Vogel, H. J. Current Understanding of Fatty Acid Biosynthesis and the Acyl Carrier Protein. *Biochem. J.* **2010**, *430*, 1–19.

(64) Cronan, J. E.; Thomas, J. Bacterial Fatty Acid Synthesis and Its Relationships with Polyketide Synthetic Pathways. *Methods Enzymol.* **2009**, *459*, 395–433.

(65) Du, L.; Sanchez, C.; Shen, B. Hybrid Peptide Polyketide Natural Products: Biosynthesis and Prospects toward Engineering Novel Molecules. *Metab. Eng.* **2001**, *3*, 78–95.

(66) Chen, H.; Du, L. Iterative Polyketide Biosynthesis by Modular Polyketide Synthases in Bacteria. *Appl. Microbiol. Biotechnol.* **2016**, *100* (2), 541–557.

(67) Hertweck, C.; Luzhetskyy, A.; Rebets, Y.; Bechthold, A. Type II Polyketide Synthases: Gaining a Deeper Insight into Enzymatic Teamwork. *Nat. Prod. Rep.* **2007**, *24*, 162–190.

(68) Butland, G.; Peregrin-Alvarez, J. M.; Li, J.; Yang, W.; Yang, X.; Canadien, V.; Starostine, A.; Richards, D.; Beattie, B.; Krogan, N.; et al. Interaction Network Containing Conserved and Essential Protein Complexes in Escherichia Coli. *Nature* **2005**, *433*, 531–537.

(69) Mercer, A. C.; Burkart, M. D. The Ubiquitous Carrier Protein-a Window to Metabolite Biosynthesis. *Nat. Prod. Rep.* **2007**, *24* (4), 750–773.

(70) Cronan, J. E. The Chain-Flipping Mechanism of ACP (Acyl Carrier Protein)-Dependent Enzymes

Appears Universal. *Biochem. J.* **2014**, *460*, 157–163.

(71) Lambalot, R. H.; Gehring, A. M.; Flugel, R. S.; Zuber, P.; LaCelle, M.; Marahiel, M. A.; Reid, R.; Khosla, C.; Walsh, C. T. A New Enzyme Superfamily - the Phosphopantetheinyl Transferases. *Chem. Biol.* **1996**, *3*, 923–936.

(72) Holak, T. A.; Kearsley, S. K.; Kim, Y.; Prestegard, J. H. Three-Dimensional Structure of Acyl Carrier Protein Determined by NMR Pseudoenergy and Distance Geometry Calculations. *Biochemistry* **1988**, *27*, 6135–6142.

(73) Qiu, X.; Janson, C. A. Structure of Apo Acyl Carrier Protein and a Proposal to Engineer Protein Crystallization through Metal Ions. *Acta Crystallogr. Sect. D* **2004**, *60*, 1545–1554.

(74) Roujeinikova, A.; Simon, W. J.; Gilroy, J.; Rice, D. W.; Rafferty, J. B.; Slabas, A. R. Structural Studies of Fatty Acyl-(Acyl Carrier Protein) Thioesters Reveal a Hydrophobic Binding Cavity That Can Expand to Fit Longer Substrates. *J. Mol. Biol.* **2007**, *365*, 135–145.

(75) Andrec, M.; Hill, R. B.; Prestegard, J. H. Amide Exchange Rates in Escherichia Coli Acyl Carrier Protein: Correlation with Protein Structure and Dynamics. *Protein Sci.* **1995**, *4*, 983–993.

(76) Xu, G.-Y.; Tam, A.; Lin, L.; Hixon, J.; Fritz, C. C.; Powers, R. Solution Structure of B. Subtilis Acyl Carrier Protein. *Structure* **2001**, *9*, 277–287.

(77) Rock, C. O.; Jackowski, S. Regulation of Phospholipid Synthesis in Escherichia Coli. *J. Biol. Chem.* **1982**, *257* (18), 10759–10765.

(78) Modak, R.; Sinha, S.; Surolia, N. Isothermal Unfolding Studies on the Apo and Holo Forms of Plasmodium Falciparum Acyl Carrier Protein. *FEBS J.* **2007**, *274*, 3313–3326.

(79) Evans, S. E.; Williams, C.; Arthur, C. J.; Burston, S. G.; Simpson, T. J.; Crosby, J.; Crump, M. P. An ACP Structural Switch: Conformational Differences between the Apo and Holo Forms of the Actinorhodin Polyketide Synthase Acyl Carrier Protein. *ChemBioChem* **2008**, *9*, 2424–2432.

(80) Rock, C. O.; Cronan, J. E.; Armitage, I. M. Molecular Properties of Acyl Carrier Protein Derivatives. *J. Bacteriol.* **1981**, *256* (6), 2669–2674.

(81) Roujeinikova, A.; Baldock, C.; Simon, W. J.; Gilroy, J.; Baker, P. J.; Stuitje, A. R.; Rice, D. W.; Slabas, A. R.; Rafferty, J. B. X-Ray Crystallographic Studies on Butyryl-ACP Reveal Flexibility of the Structure around a Putative Acyl Chain Binding Site. *Structure* **2002**, *10*, 825–835.

(82) Ploskon, E.; Arthur, C. J.; Kanari, A. L. P.; Wattana-amorn, P.; Williams, C.; Crosby, J.; Simpson,

T. J.; Willis, C. L.; Crump, M. P. Recognition of Intermediate Functionality by Acyl Carrier Protein over a Complete Cycle of Fatty Acid Biosynthesis. *Chem. Biol.* **2010**, *17*, 776–785.

(83) Beld, J.; Cang, H.; Burkart, M. D. Visualizing the Chain-Flipping Mechanism in Fatty-Acid Biosynthesis. *Angew. Chemie - Int. Ed.* **2014**, *53*, 14456–14461.

(84) Zhang, Y.-M.; Marrakchi, H.; White, S. W.; Rock, C. O. The Application of Computational Methods to Explore the Diversity and Structure of Bacterial Fatty Acid Synthase. *J. Lipid Res.* **2003**, *44*, 1–10.

(85) Parris, K. D.; Lin, L.; Tam, A.; Mathew, R.; Hixon, J.; Stahl, M.; Fritz, C. C.; Seehra, J.; Somers, W. S. Crystal Structures of Substrate Binding to Bacillus Subtilis Holo-(Acyl Carrier Protein) Synthase Reveal a Novel Trimeric Arrangement of Molecules Resulting in Three Active Sites. *Structure* **2000**, *8* (8), 883–895.

(86) Alekseyev, V. Y.; Liu, C. W.; Cane, D. E.; Puglisi, J. D.; Khosla, C. Solution Structure and Proposed Domain – Domain Recognition Interface of an Acyl Carrier Protein Domain from a Modular Polyketide Synthase. *Protein Sci.* **2007**, *16*, 2093–2107.

(87) Ploskon, E.; Arthur, C. J.; Evans, S. E.; Williams, C.; Crosby, J.; Simpson, T. J.; Crump, M. P. A Mammalian Type I Fatty Acid Synthase Acyl Carrier Protein Domain Does Not Sequester Acyl Chains. *J. Biol. Chem.* **2008**, *283* (1), 518–528.

(88) Beld, J.; Sonnenschein, E. C.; Vickery, C. R.; Noel, J. P.; Burkart, M. D. The Phosphopantetheinyl Transferases: Catalysis of a Post-Translational Modification Crucial for Life. *Nat. Prod. Rep.* **2014**, *31*, 61–108.

(89) Leibundgut, M.; Maier, T.; Jenni, S.; Ban, N. The Multienzyme Architecture of Eukaryotic Fatty Acid Synthases. *Curr. Opin. Struct. Biol.* **2008**, *18*, 714–725.

(90) Elovson, J.; Vagelos, P. R. Acyl Carrier Protein X. Acyl Carrier Protein Synthetase. *J. Biol. Chem.* **1968**, *243* (13), 3603–3611.

(91) Gehring, A. M.; Lambalot, R. H.; Vogel, K. W.; Drueckhammer, D. G.; Walsh, C. T. Ability of Streptomyces Spp. Acyl Carrier Proteins and Coenzyme A Analogs to Serve as Substrates in Vitro for E. Coli Holo-ACP Synthase. *Chem. Biol.* **1997**, *4* (1), 17–24.

(92) Reuter, K.; Mofid, M. R.; Marahiel, M. A.; Ficner, R. Crystal Structure of the Surfactin Synthetase-Activating Enzyme Sfp: A Prototype of the 4'-phosphopantetheinyl Transferase Superfamily. *EMBO J.* **1999**, *18* (23), 6823–6831.

(93) Raman, M. C. C.; Johnson, K. A.; Clarke, D. J.; Naismith, J. H.; Campopiano, D. J. The Serine Palmitoyltransferase from Sphingomonas Wittichii RW1: An Interesting Link to an Unusual Acyl Carrier Protein. *Biopolymers* **2010**, *93* (9), 811–822.

(94) Trivedi, O. A.; Arora, P.; Sridharan, V.; Tickoo, R.; Mohanty, D.; Gokhale, R. S. Enzymic Activation and Transfer of Fatty Acids as Acyl-Adenylates in Mycobacteria. *Nature* **2004**, *428*, 441–444.

(95) Kosa, N. M.; Haushalter, R. W.; Smith, A. R.; Burkart, M. D. Reversible Labeling of Native and Fusion-Protein Motifs. *Nat. Methods* **2012**, *9* (10), 981–987.

(96) Huang, W.; Jia, J.; Edwards, P.; Dehesh, K.; Schneider, G.; Lindqvist, Y. Crystal Structure of β-Ketoacyl-Acyl Carrier Protein Synthase II from E. Coli Reveals the Molecular Architecture of Condensing Enzymes. *EMBO J.* **1998**, *17* (5), 1183–1191.

(97) Robbins, T.; Kapilivsky, J.; Cane, D. E.; Khosla, C. Roles of Conserved Active Site Residues in the Ketosynthase Domain of an Assembly Line Polyketide Synthase. *Biochemistry* **2016**, *55*, 4476–4484.

(98) Das, A.; Khosla, C. Biosynthesis of Aromatic Polyketides in Bacteria. *Acc. Chem. Res.* **2009**, *42* (5), 631–639.

(99) Bao, W.; Wendt-Pienkowski, E.; Hutchinson, C. R. Reconstitution of the Iterative Type II Polyketide Synthase for Tetracenomycin F2 Biosynthesis. *Biochemistry* **1998**, *37*, 8132–8138.

(100) McDaniel, R.; Ebert-Khosla, S.; Hopwood, D. A.; Khosla, C. Engineered Biosynthesis of Novel Polyketides. *Science* **1993**, *262*, 1546–1550.

(101) Nicholson, T. P.; Winfield, C.; Westcott, J.; Crosby, J.; Simpson, T. J.; Cox, R. J. First in Vitro Directed Biosynthesis of New Compounds by a Minimal Type II Polyketide Synthase: Evidence for the Mechanism of Chain Length Determination. *Chem. Commun.* **2003**, 686–687.

(102) Keatinge-Clay, A. T.; Maltby, D. A.; Medzihradszky, K. F.; Khosla, C.; Stroud, R. M. An Antibiotic Factory Caught in Action. *Nat. Struct. Mol. Biol.* **2004**, *11* (9), 888–893.

(103) Tang, Y.; Tsai, S.-C.; Khosla, C. Polyketide Chain Length Control by Chain Length Factor. *J. Am. Chem. Soc.* **2003**, *125* (42), 12708–12709.

(104) Bisang, C.; Long, P. F.; Cortes, J.; Westcott, J.; Crosby, J.; Matharu, A.-L.; Cox, R. J.; Simpson, T. J.; Staunton, J.; Leadlay, P. F. A Chain Initiation Factor Common to Both Modular and Aromatic Polyketide Synthases. *Nature* **1999**, *401*, 502–505.

(105) Witkowski, A.; Joshi, A. K.; Lindqvist, Y.; Smith, S. Conversion of a β-Ketoacyl Synthase to a Malonyl Decarboxylase by Replacement of the Active-Site Cysteine with Glutamine. *Biochemistry* **1999**, *38*, 11643–11650.

(106) Fernandez-Moreno, M. A.; Martinez, E.; Botosli, L.; Hopwood, D. A.; Malpartida, F. Nucleotide Sequence and Deduced Functions of a Set of Cotranscribed Genes of Streptomyces Coelicolor A3(2) Including the Polyketide Synthase for the Antibiotic Actinorhodin. *J. Biol. Chem.* **1992**, *27*, 19278–19290.

(107) Carreras, C. W.; Khosla, C. Purification and in Vitro Reconstitution of the Essential Protein Components of an Aromatic Polyketide Synthase. *Biochemistry* **1998**, *37*, 2084–2088.

(108) Berg, P. Acyl Adenylates: An Enzymatic Mechanism of Acetate Activation. *J. Biol. Chem.* **1956**, *222*, 991–1013.

(109) Schmelz, S.; Naismith, J. H. Adenylate-Forming Enzymes. *Curr. Opin. Struct. Biol.* **2009**, *19* (6), 666–671.

(110) Gulick, A. M. Conformational Dynamics in the Acyl-CoA Synthetases, Adenylation Domains of Non-Ribosomal Peptide Synthetases, and Firefly Luciferase. *ACS Chem. Biol.* **2009**, *4* (10), 811–827.

(111) Wang, M.; Moynié, L.; Harrison, P. J.; Kelly, V.; Piper, A.; Naismith, J. H.; Campopiano, D. J. Using the Pimeloyl-CoA Synthetase Adenylation Fold to Synthesize Fatty Acid Thioesters. *Nat. Chem. Biol.* **2017**, *13* (6), 660–667.

(112) Estrada, P.; Manandhar, M.; Dong, S.-H.; Deveryshetty, J.; Agarwal, V.; Cronan, J. E.; Nair, S. K. The Pimeloyl-CoA Synthetase BioW Defines a New Fold for Adenylate-Forming Enzymes. *Nat. Chem. Biol.* **2017**, *13*, 668–674.

(113) Conti, E.; Franks, N. P.; Brick, P. Crystal Structure of Firefly Luciferase Throws Light on a Superfamily of Adenylate-Forming Enzymes. *Structure* **1996**, *4* (3), 287–298.

(114) McElroy, W. D.; DeLuca, M.; Travis, J. Molecular Uniformity in Biological Catalyses. *Science* **1967**, *157* (3785), 150–160.

(115) Gocht, M.; Marahiel, M. A. Analysis of Core Sequences in the D-Phe Activating Domain of the Multifunctional Peptide Synthetase TycA by Site-Directed Mutagenesis. *J. Bacteriol.* **1994**, *176* (9), 2654–2662.

(116) Kochan, G.; Pilka, E. S.; von Delft, F.; Oppermann, U.; Yue, W. W. Structural Snapshots for the

Conformation-Dependent Catalysis by Human Medium-Chain Acyl-Coenzyme A Synthetase ACSM2A. *J. Mol. Biol.* **2009**, *388* (5), 997–1008.

(117) Reger, A. S.; Wu, R.; Dunaway-Mariano, D.; Gulick, A. M. Structural Characterization of a 140° Domain Movement in the Two-Step Reaction Catalyzed by 4-chlorobenzoate:CoA Ligase. *Biochemistry* **2008**, *47*, 8016–8025.

(118) Wu, R.; Cao, J.; Lu, X.; Reger, A. S.; Gulick, A. M.; Dunway-Mariano, D. Mechanism of 4-Chlorobenzoate: Coenzyme A Ligase Catalysis. *Biochemistry* **2013**, *47*, 8026–8039.

(119) Du, L.; Lou, L. PKS and NRPS Release Mechanisms. *Nat. Prod. Rep.* **2010**, *27*, 255–278.

(120) Mullowney, M. W.; McClure, R. A.; Robey, M. T.; Kelleher, N. L.; Thomson, R. J. Natural Products from Thioester Reductase Containing Biosynthetic Pathways. *Nat. Prod. Rep.* **2018**, *Advance Ar*.

(121) Awodi, U. R.; Ronan, J. L.; Masschelein, J.; de los Santos, E. L. C.; Challis, G. L. Thioester Reduction and Aldehyde Transamination Are Universal Steps in Actinobacterial Polyketide Alkaloid Biosynthesis. *Chem. Sci.* **2017**, *8* (1), 411–415.

(122) Chhabra, A.; Haque, A. S.; Kant Pal, R.; Goyal, A.; Rai, R.; Joshi, S.; Panjikar, S.; Pasha, S.; Sankaranarayanan, R.; Gokhale, R. S. Nonprocessive [2+2]e− off-Loading Reductase Domains from Mycobacterial Nonribosomal Peptide Synthetases. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (15), 5681–5686.

(123) Ehmann, D. E.; Gehring, A. M.; Walsh, C. T. Lysine Biosynthesis in Saccharomyces Cerevisiae: Mechanism of α-Aminoadipate Reductase (Lys2) Involves Posttranslational Phosphopantetheinylation by Lys5. *Biochemistry* **1999**, *38*, 6171–6177.

(124) Barajas, J. F.; Phelan, R. M.; Schaub, A. J.; Kliewer, J. T.; Kelly, P. J.; Jackson, D. R.; Luo, R.; Keasling, J. D.; Tsai, S.-C. Comprehensive Structural and Biochemical Analysis of the Terminal Myxalamid Reductase Domain for the Engineered Production of Primary Alcohols. *Chem. Biol.* **2015**, *22*, 1018–1029.

(125) Gahloth, D.; Dunstan, M. S.; Quaglia, D.; Klumbys, E.; Lockhart-Cairns, M. P.; Hill, A. M.; Derrington, S. R.; Scrutton, N. S.; Turner, N. J.; Leys, D. Structures of Carboxylic Acid Reductase Reveal Domain Dynamics Underlying Catalysis. *Nat. Chem. Biol.* **2017**, *13*, 975–981.

(126) Lin, F.; Das, D.; Lin, X. N.; Marsh, E. N. G. Aldehyde-Forming Fatty Acyl-CoA Reductase from Cyanobacteria: Expression, Purification and Characterization of the Recombinant Enzyme.

FEBS J. **2013**, *280*, 4773–4781.

(127) Warui, D. M.; Pandelia, M.-E.; Rajakovich, L. J.; Krebs, C.; Bollinger Jr., J. M.; Booker, S. J. Efficient Delivery of Long-Chain Fatty Aldehydes from the Nostoc Punctiforme Acyl−Acyl Carrier Protein Reductase to Its Cognate Aldehyde-Deformylating Oxygenase. *Biochemistry* **2015**, *54*, 1006–1015.

(128) Gaitatzis, N.; Kunze, B.; Mueller, R. In Vitro Reconstitution of the Myxochelin Biosynthetic Machinery of Stigmatella Aurantiaca Sg a15: Biochemical Characterization of a Reductive Release Mechanism from Nonribosomal Peptide Synthetases. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98* (20), 11136–11141.

(129) Read, J. A.; Walsh, C. T. The Lyngbyatoxin Biosynthetic Assembly Line: Chain Release by Four-Electron Reduction of a Dipeptidyl Thioester to the Corresponding Alcohol. *J. Am. Chem. Soc.* **2007**, *129*, 15762–15763.

(130) Li, Y.; Weissman, K. J.; Mueller, R. Myxochelin Biosynthesis: Direct Evidence for Two- and Four-Electron Reduction of a Carrier Protein-Bound Thioester. *J. Am. Chem. Soc.* **2008**, *130*, 7554–7555.

(131) Heyl, D.; Harris, S. A.; Folkers, K. The Chemistry of Vitamin B6. VI. Pyridoxylamino Acids. *J. Am. Chem. Soc.* **1948**, *70* (10), 3429–3431.

(132) Jansonius, J. N. Structure, Evolution and Action of Vitamin B6-Dependent Enzymes. *Curr. Opin. Struct. Biol.* **1998**, *8*, 759–769.

(133) Steffen-Munsberg, F.; Vickers, C.; Kohls, H.; Land, H.; Mallin, H.; Nobili, A.; Skalden, L.; van den Bergh, T.; Joosten, H.-J.; Berglund, P.; et al. Bioinformatic Analysis of a PLP-Dependent Enzyme Superfamily Suitable for Biocatalytic Applications. *Biotechnol. Adv.* **2015**, *33*, 566–604.

(134) Schneider, G.; Käck, H.; Lindqvist, Y. The Manifold of Vitamin B6 Dependent Enzymes. *Structure* **2000**, *8* (1), 1–6.

(135) Eliot, A. C.; Kirsch, J. F. Pyridoxal Phosphate Enzymes: Mechanistic, Structural, and Evolutionary Considerations. *Annu. Rev. Biochem.* **2004**, *73*, 383–415.

(136) Dunathan, H. C. Conformation and Reaction Specificity in Pyridoxal Phosphate Enzymes. *Proc. Natl. Acad. Sci. U. S. A.* **1966**, *55*, 712–716.

(137) Ford, G. C.; Eichele, G.; Jansonius, J. N. Three-Dimensional Structure of a Pyridoxal-Phosphate-Dependent Enzyme, Mitochondrial Aspartate Aminotransferase. *Proc. Natl. Acad. Sci. U. S. A.*

1980, *77* (5), 2559–2563.

(138) Hyde, C. C.; Ahmed, S. A.; Padlan, E. A.; Miles, E. W.; Davies, D. R. Three-Dimensional Structure of the Tryptophan Synthase α2β2 Multienzyme Complex from Salmonella Typhimurium. *J. Biol. Chem.* **1988**, *263* (33), 17857–17871.

(139) Sugio, S.; Petsko, G. A.; Manning, J. M.; Soda, K.; Ringe, D. Crystal Structure of D-Amino Acid Aminotransferase: How the Protein Controls Stereoselectivity. *Biochemistry* **1995**, *34*, 9661–9669.

(140) Shaw, J. P.; Petsko, G. A.; Ringe, D. Determination of the Structure of Alanine Racemase from Bacillus Stearothermophilus at 1.9Å Resolution. *Biochemistry* **1997**, *36*, 1329–1342.

(141) Hunter, G. A.; Ferreira, G. C. Molecular Enzymology of 5-Aminolevulinate Synthase, the Gatekeeper of Heme Biosynthesis. *Biochim. Biophys. Acta* **2011**, *1814*, 1467–1473.

(142) Webster, S. P.; Alexeev, D.; Campopiano, D. J.; Watt, R. M.; Alexeeva, M.; Sawyer, L.; Baxter, R. L. Mechanism of 8-Amino-7-Oxononanoate Synthase: Spectroscopic, Kinetic, and Crystallographic Studies. *Biochemistry* **2000**, *39*, 516–528.

(143) Lowther, J.; Naismith, J. H.; Dunn, T. M.; Campopiano, D. J. Structural, Mechanistic and Regulatory Studies of Serine Palmitoyltransferase. *Biochem. Soc. Trans.* **2012**, *40*, 547–554.

(144) Harrison, P. J.; Dunn, T. M.; Campopiano, D. J. Sphingolipid Biosynthesis in Man and Microbes. *Nat. Prod. Rep.* **2018**, *Advance Ar*.

(145) Kellmann, R.; Mihali, T. K.; Young, J. J.; Pickford, R.; Pomati, F.; Neilan, B. A. Biosynthetic Intermediate Analysis and Functional Homology Reveal a Saxitoxin Gene Cluster in Cyanobacteria. *Appl. Environ. Microbiol.* **2008**, *74* (13), 4044–4053.

(146) Gerber, R.; Lou, L.; Du, L. A PLP-Dependent Polyketide Chain Releasing Mechanism in the Biosynthesis of Mycotoxin Fumonisins in Fusarium Verticillioides. *J. Am. Chem. Soc.* **2009**, *131*, 3148–3149.

(147) Zhang, J.; Ferreira, G. C. Transient State Kinetic Investigation of 5-Aminolevulinate Synthase Reaction Mechanism. *J. Biol. Chem.* **2002**, *277* (47), 44660–44669.

(148) Ploux, O.; Marquet, A. Mechanistic Studies on the 8-Amino-7-Oxopelargonate Synthase, a Pyridoxal-5'-phosphate-Dependent Enzyme Involved in Biotin Biosynthesis. *Eur. J. Biochem.* **1996**, *236*, 301–308.

(149) Kerbarh, O.; Campopiano, D. J.; Baxter, R. L. Mechanism of α-Oxoamine Synthases: Identification of the Intermediate Claisen Product in the 8-Amino-7-Oxononanoate Synthase Reaction. *Chem. Commun.* **2006**, *1*, 60–62.

(150) Bashir, Q.; Rashid, N.; Akhtar, M. Mechanism and Substrate Stereochemistry of 2-Amino-3-Oxobutyrate CoA Ligase: Implications for 5-Aminolevulinate Synthase and Related Enzymes. *Chem. Commun.* **2006**, *48*, 5065–5067.

(151) Yard, B. A.; Carter, L. G.; Johnson, K. A.; Overton, I. M.; Dorward, M.; Liu, H.; McMahon, S. A.; Oke, M.; Puech, D.; Barton, G. J.; et al. The Structure of Serine Palmitoyltransferase; Gateway to Sphingolipid Biosynthesis. *J. Mol. Biol.* **2007**, *370*, 870–886.

(152) Ikushiro, H.; Islam, M. M.; Okamoto, A.; Hoseki, J.; Murakawa, T.; Fujii, S.; Miyahara, I.; Hayashi, H. Structural Insights into the Enzymatic Mechanism of Serine Palmitoyltransferase from Sphingobacterium Multivorum. *J. Biochem.* **2009**, *146* (4), 549–562.

(153) Astner, I.; Schulze, J. O.; van den Heuvel, J.; Jahn, D.; Schubert, W.-D.; Heinz, D. W. Crystal Structure of 5-Aminolevulinate Synthase, the First Enzyme of Heme Biosynthesis, and Its Link to XLSA in Humans. *EMBO J.* **2005**, *24*, 3166–3177.

(154) Franklin, M. C.; Cheung, J.; Rudolph, M. J.; Burshteyn, F.; Cassidy, M.; Gary, E.; Hillerich, B.; Yao, Z. K.; Carlier, P. R.; Totrov, M.; et al. Structural Genomics for Drug Design against the Pathogen Coxiella Burnetii. *Proteins* **2015**, *83* (12), 2124–2136.

(155) Schmidt, A.; Sivaraman, J.; Li, Y.; Larocque, R.; Barbosa, J. A.; Smith, C.; Matte, A.; Schrag, J. D.; Cygler, M. Three-Dimensional Structure of 2-Amino-3-Ketobutyrate CoA Ligase from Escherichia Coli Complexed with a PLP-Substrate Intermediate: Inferred Reaction Mechanism. *Biochemistry* **2001**, *40* (17), 5151–5160.

(156) Jahan, N.; Potter, J. A.; Sheikh, M. A.; Botting, C. H.; Shirran, S. L.; Westwood, N. J.; Taylor, G. L. Insights into the Biosynthesis of the Vibrio Cholerae Major Autoinducer CAI-1 from the Crystal Structure of the PLP-Dependent Enzyme CqsA. *J. Mol. Biol.* **2009**, *392*, 763–773.

(157) Ikushiro, H.; Hayashi, H. Mechanistic Enzymology of Serine Palmitoyltransferase. *Biochim. Biophys. Acta* **2011**, *1814*, 1474–1480.

(158) Mehta, P. K.; Hale, T. I.; Christen, P. Aminotransferases: Demonstration of Homology and Division into Evolutionary Subgroups. *Eur. J. Biochem.* **1993**, *214*, 549–561.

(159) Slabu, I.; Galman, J. L.; Lloyd, R. C.; Turner, N. J. Discovery, Engineering, and Synthetic

Application of Transaminase Biocatalysts. *ACS Catal.* **2017**, *7*, 8263–8284.

(160) Oshima, T.; Tamiya, N. Mechanism of Transaminase Action. *Biochem. J.* **1961**, *78*, 116–119.

(161) Velick, S. F.; Vavra, J. A Kinetic and Equilibrium Analysis of the Glutamic Oxaloacetate Transaminase Mechanism. *J. Biol. Chem.* **1962**, *237* (7), 2109–2122.

(162) Cassimjee, K. E.; Manta, B.; Himo, F. A Quantum Chemical Study of the ω-Transaminase Reaction Mechanism. *Org. Biomol. Chem.* **2015**, *13*, 8453–8464.

(163) Yoshimura, T.; Jhee, K.-H.; Soda, K. Stereospecificity for the Hydrogen Transfer and Molecular Evolution of Pyridoxal Enzymes. *Biosci. Biotechnol. Biochem.* **1996**, *60* (2), 181–187.

(164) Shen, B. W.; Hennig, M.; Hohenester, E.; Jansonius, J. N.; Schirmer, T. Crystal Structure of Human Recombinant Ornithine Aminotransferase. *J. Mol. Biol.* **1998**, *277* (1), 81–102.

(165) Liu, W.; Peterson, P. E.; Carter, R. J.; Zhou, X.; Langston, J. A.; Fisher, A. J.; Toney, M. D. Crystal Structures of Unbound and Aminooxyacetate-Bound Escherichia Coli γ-Aminobutyrate Aminotransferase. *Biochemistry* **2004**, *43*, 10896–10905.

(166) Zhu, Y.; Xu, J.; Mei, X.; Feng, Z.; Zhang, L.; Zhang, Q.; Zhang, G.; Zhu, W.; Liu, J.; Zhang, C. Biochemical and Structural Insights into the Aminotransferase CrmG in Caerulomycin Biosynthesis. *ACS Chem. Biol.* **2016**, *11* (4), 943–952.

(167) Cha, H. J.; Jeong, J.-H.; Rojviriya, C.; Kim, Y.-G. Structure of Putrescine Aminotransferase from Escherichia Coli Provides Insights into the Substrate Specificity among Class III Aminotransferases. *PLoS One* **2014**, *9* (11), 1–15.

(168) Thomas, M. G.; Burkart, M. D.; Walsh, C. T. Conversion of L-Proline to Pyrrolyl-2-Carboxyl- S -PCP during Undecylprodigiosin and Pyoluteorin Biosynthesis. *Chem. Biol.* **2002**, *9*, 171–184.

(169) Franke, J.; Hertweck, C. Biomimetic Thioesters as Probes for Enzymatic Assembly Lines: Synthesis, Applications, and Challenges. *Cell Chem. Biol.* **2016**, *23* (10), 1179–1192.

(170) Beld, J.; Lee, D. J.; Burkart, M. D. Fatty Acid Biosynthesis Revisited: Structure Elucidation and Metabolic Engineering. *Mol. Biosyst.* **2015**, *11*, 38–59.

(171) Robert, X.; Gouet, P. Deciphering Key Features in Protein Structures with the New ENDscript Server. *Nucleic Acids Res.* **2014**, *42*, 320–324.

(172) Herbst, D. A.; Jakob, R. P.; Zähringer, F.; Maier, T. Mycocerosic Acid Synthase Exemplifies the Architecture of Reducing Polyketide Synthases. *Nature* **2016**, *531*, 533–537.

(173) Whicher, J. R.; Smaga, S. S.; Hansen, D. A.; Brown, W. C.; Gerwick, W. H.; Sherman, D. H.; Smith, J. L. Cyanobacterial Polyketide Synthase Docking Domains: A Tool for Engineering Natural Product Biosynthesis. *Chem. Biol.* **2013**, *20*, 1340–1351.

(174) Gay, D. C.; Wagner, D. T.; Meinke, J. L.; Zogzas, C. E.; Gay, G. R.; Keatinge-Clay, A. T. The LINKS Motif Zippers Trans-Acyltransferase Polyketide Synthase Assembly Lines into a Biosynthetic Megacomplex. *J. Struct. Biol.* **2016**, *193*, 196–205.

(175) Gay, D. C.; Gay, G.; Axelrod, A. J.; Jenner, M.; Kohlhaas, C.; Kampa, A.; Oldham, N. J.; Piel, J.; Keatinge-Clay, A. T. A Close Look at a Ketosynthase from a Trans-Acyltransferase Modular Polyketide Synthase. *Structure* **2014**, *22* (3), 444–451.

(176) Sucipto, H.; Sahner, J. H.; Prusov, E.; Wenzel, S. C.; Hartmann, R.; Koehnke, J.; Mueller, R. In Vitro Reconstitution of α-Pyrone Ring Formation in Myxopyronin Biosynthesis. *Chem. Sci.* **2015**, *6*, 5076–5085.

(177) Von Wettstein-Knowles, P.; Olsen, J. G.; McGuire, K. A.; Henriksen, A. Fatty Acid Synthesis: Role of Active Site Histidines and Lysine in Cys-His-His-Type Beta-Ketoacyl-Acyl Carrier Protein Synthases. *FEBS J.* **2006**, *273* (4), 695–710.

(178) Kusebauch, B.; Brendel, N.; Kirchner, H.; Dahse, H.-M.; Hertweck, C. Assessing Oxazole Bioisosteres as Mutasynthons on the Rhizoxin Assembly Line. *ChemBioChem* **2011**, *12* (15), 2284–2288.

(179) Mann, S.; Ploux, O. Pyridoxal-5'-Phosphate-Dependent Enzymes Involved in Biotin Biosynthesis: Structure, Reaction Mechanism and Inhibition. *Biochim. Biophys. Acta - Proteins Proteomics* **2011**, *1814* (11), 1459–1465.

(180) Alexeev, D.; Alexeeva, M.; Baxter, R. L.; Campopiano, D. J.; Webster, S. P.; Sawyer, L. The Crystal Structure of 8-Amino-7-Oxononanoate Synthase: A Bacterial PLP-Dependent, Acyl-CoA-Condensing Enzyme. *J. Mol. Biol.* **1998**, *284* (2), 401–419.

(181) Liu, H.; Naismith, J. H. A Simple and Efficient Expression and Purification System Using Two Newly Constructed Vectors. *Protein Expr. Purif.* **2009**, *63* (2), 102–111.

(182) Quadri, L. E. N.; Weinreb, P. H.; Lei, M.; Nakano, M. M.; Zuber, P.; Walsh, C. T. Characterization of Sfp, a Bacillus Subtilis Phosphopantetheinyl Transferase for Peptidyl Carrier Protein Domains in Peptide Synthetases. *Biochemistry* **1998**, *37*, 1585–1595.

(183) Hitchman, T. S.; Crosby, J.; Byrom, K. J.; Cox, R. J.; Simpson, T. J. Catalytic Self-Acylation of Type

II Polyketide Synthase Acyl Carrier Proteins. *Chem. Biol.* **1998**, *5* (1), 35–47.

(184) Das, A.; Khosla, C. Biosynthesis of Aromatic Polyketides in Bacteria. *Acc. Chem. Res.* **2009**, *42* (5), 631–639.

(185) Bhor, V. M.; Dev, S.; Vasanthakumar, G. R.; Kumar, P.; Sinha, S.; Surolia, A. Broad Substrate Stereospecificity of the Mycobacterium Tuberculosis 7-Keto-8-Aminopelargonic Acid Synthase: Spectroscopic and Kinetic Studies. *J. Biol. Chem.* **2006**, *281* (35), 25076–25088.

(186) Ikushiro, H.; Fujii, S.; Shiraiwa, Y.; Hayashi, H. Acceleration of the Substrate Cα Deprotonation by an Analogue of the Second Substrate Palmitoyl-CoA in Serine Palmitoyltransferase. *J. Biol. Chem.* **2008**, *283* (12), 7542–7553.

(187) Paik, P. K.; Rudin, C. M.; Pietanza, M. C.; Brown, A.; Rizvi, N. A.; Takebe, N.; Travis, W.; James, L.; Ginsberg, M. S.; Juergens, R.; et al. A Phase II Study of Obatoclax Mesylate, a Bcl-2 Antagonist, plus Topotecan in Relapsed Small Cell Lung Cancer. *Lung Cancer* **2011**, *74* (3), 481–485.

(188) Lerchner, A.; Daake, M.; Jarasch, A.; Skerra, A. Fusion of an Alcohol Dehydrogenase with an Aminotransferase Using a PAS Linker to Improve Coupled Enzymatic Alcohol-to-Amine Conversion. *Protein Eng. Des. Sel.* **2016**, *29* (12), 557–562.

(189) Ellis, J. M.; Frahm, J. L.; Li, L. O.; Coleman, R. A. Acyl-Coenzyme A Synthetases in Metabolic Control. *Curr. Opin. Lipidol.* **2010**, *21* (3), 212–217.

(190) Zhang, Z.; Zhou, R.; Sauder, J. M.; Tonge, P. J.; Burley, S. K.; Swaminathan, S. Structural and Functional Studies of Fatty Acyl Adenylate Ligases from E. Coli and L. Pneumophila. *J. Mol. Biol.* **2011**, *406* (2), 313–324.

(191) Li, W.; Gu, S.; Fleming, J.; Bi, L. Crystal Structure of FadD32, an Enzyme Essential for Mycolic Acid Biosynthesis in Mycobacteria. *Sci. Rep.* **2015**, *5*, 15493.

(192) Kuhn, M. L.; Alexander, E.; Minasov, G.; Page, H. J.; Wawrzak, Z.; Shuvalova, L.; Flores, K. J.; Wilson, D. J.; Shi, C.; Aldrich, C. C.; et al. Structure of the Essential Mtb FadD32 Enzyme: A Promising Drug Target for Treating Tuberculosis. *ACS Infect. Dis.* **2016**, *2*, 579–591.

(193) Guillet, V.; Galandrin, S.; Maveyraud, L.; Ladevèze, S.; Mariaule, V.; Bon, C.; Eynard, N.; Daffé, M.; Marrakchi, H.; Mourey, L. Insight into Structure-Function Relationships and Inhibition of the Fatty Acyl-AMP Ligase (FadD32) Orthologs from Mycobacteria. *J. Biol. Chem.* **2016**, *291* (15), 7973–7989.

(194) Paul, S.; Ishida, H.; Nguyen, L. T.; Liu, Z.; Vogel, H. J. Structural and Dynamic Characterization of a Freestanding Acyl Carrier Protein Involved in the Biosynthesis of Cyclic Lipopeptide Antibiotics. *Protein Sci.* **2017**, *26*, 946–959.

(195) Wattana-amorn, P.; Williams, C.; Płoskoń, E.; Cox, R. J.; Simpson, T. J.; Crosby, J.; Crump, M. P. Solution Structure of an Acyl Carrier Protein Domain from a Fungal Type I Polyketide Synthase. *Biochemistry* **2010**, *49*, 2186–2193.

(196) Lössl, P.; Snijder, J.; Heck, A. J. R. Boundaries of Mass Resolution in Native Mass Spectrometry. *Am. Soc. Mass Spectrom.* **2014**, *25*, 906–917.

(197) Pinkerton, D. M.; Banwell, M. G.; Willis, A. C. Total Syntheses of Tambjamines C, E, F, G, H, I and J, BE-18591, and a Related Alkaloid from the Marine Bacterium *Pseudoalteromonas Tunicata*. *Org. Lett.* **2007**, *24*, 27–30.

(198) Wlodawer, A.; Minor, W.; Dauter, Z.; Jaskolski, M. Protein Crystallography for Non-Crystallographers, or How to Get the Best (but Not More) from Published Macromolecular Structures. *FEBS J.* **2008**, *275* (1), 1–21.

(199) Miller, B. R.; Drake, E. J.; Shi, C.; Aldrich, C. C.; Gulick, A. M. Structures of a Nonribosomal Peptide Synthetase Module Bound to MbtH-like Proteins Support a Highly Dynamic Domain Architecture. *J. Biol. Chem.* **2016**, *291* (43), 22559–22571.

(200) Drake, E. J.; Miller, B. R.; Shi, C.; Tarrasch, J. T.; Sundlov, J. A.; Allen, C. L.; Skiniotis, G.; Aldrich, C. C.; Gulick, A. M. Structures of Two Distinct Conformations of Holo-Non-Ribosomal Peptide Synthetases. *Nature* **2016**, *529*, 235–238.

(201) Tarry, M. J.; Haque, A. S.; Bui, K. H.; Schmeing, T. M. X-Ray Crystallography and Electron Microscopy of Cross- and Multi-Module Nonribosomal Peptide Synthetase Proteins Reveal a Flexible Architecture. *Structure* **2017**, *25*, 783–793.

(202) Tanovic, A.; Samel, S. A.; Essen, L.-O.; Marahiel, M. A. Crystal Structure of the Termination Module of a Nonribosomal Peptide Synthase. *Science* **2008**, *321*, 659–664.

(203) Kelley, L. A.; Mezulis, S.; Yates, C. M.; Wass, M. N.; Sternberg, M. J. E. The Phyre2 Web Portal for Protein Modeling, Prediction and Analysis. *Nat. Protoc.* **2015**, *10* (6), 845–858.

(204) Crawford, J. M.; Thomas, P. M.; Scheerer, J. R.; Vagstad, A. L.; Kelleher, L.; Townsend, C. A.; Jason, M.; Scheerer, R.; Vagstad, L.; Kelleher, N. L.; et al. Deconstruction Polyketide of Iterative Multidomain Function Synthase. *Science* **2008**, *320* (5873), 243–246.

(205) Fellers, R. T.; Greer, J. B.; Early, B. P.; Yu, X.; LeDuc, R. D.; Kelleher, N. L.; Thomas, P. M. ProSight Lite: Graphical Software to Analyze Top-down Mass Spectrometry Data. *Proteomics* **2015**, *15* (7), 1235–1238.

(206) Williamson, N. R.; Simonsen, H. T.; Ahmed, R. A. A.; Goldet, G.; Slater, H.; Woodley, L.; Leeper, F. J.; Salmond, G. P. C. Biosynthesis of the Red Antibiotic, Prodigiosin, in Serratia: Identification of a Novel 2-Methyl-3-N-Amyl-Pyrrole (MAP) Assembly Pathway, Definition of the Terminal Condensing Enzyme, and Implications for Undecylprodigiosin Biosynthesis in Streptomyces. *Mol. Microbiol.* **2005**, *56* (4), 971–989.

(207) Lou, X.; Ran, T.; Han, N.; Gao, Y.; He, J.; Tang, L.; Xu, D.; Wang, W. Crystal Structure of the Catalytic Domain of PigE: A Transaminase Involved in the Biosynthesis of 2-Methyl-3-N-Amyl-Pyrrole (MAP) from Serratia Sp. FS14. *Biochem. Biophys. Res. Commun.* **2014**, *447* (1), 178–183.

(208) Prasad, S.; Khadatare, P. B.; Roy, I. Effect of Chemical Chaperones in Improving the Solubility of Recombinant Proteins in Escherichia Coli. *Appl. Environ. Microbiol.* **2011**, *77* (13), 4603–4609.

(209) DeSouza, L. V.; Siu, K. W. M. Mass Spectrometry-Based Quantification. *Clin. Biochem.* **2013**, *46*, 421–431.

(210) Wood, A. J. L.; Weise, N. J.; Frampton, J. D.; Dunstan, M. S.; Hollas, M. A.; Derrington, S. R.; Lloyd, R. C.; Quaglia, D.; Parmeggiani, F.; Leys, D.; et al. Adenylation Activity of Carboxylic Acid Reductases Enables the Synthesis of Amides. *Angew. Chemie - Int. Ed.* **2017**, *56*, 14498–14501.

(211) Johannes, T.; Simurdiak, M. R.; Zhao, H. Biocatalysis. *Encycl. Chem. Process.* **2006**, 101–110.

(212) Tørring, T.; Shames, S. R.; Cho, W.; Roy, C. R.; Crawford, J. M. Acyl Histidines: New N-Acyl Amides from Legionella Pneumophila. *ChemBioChem* **2017**, *18*, 638–646.

(213) Hesselink, J. M. K.; Hekker, T. A. M. Therapeutic Utility of Palmitoylethanolamide in the Treatment of Neuropathic Pain Associated with Various Pathological Conditions: A Case Series. *J. Pain Res.* **2012**, *5*, 437–442.

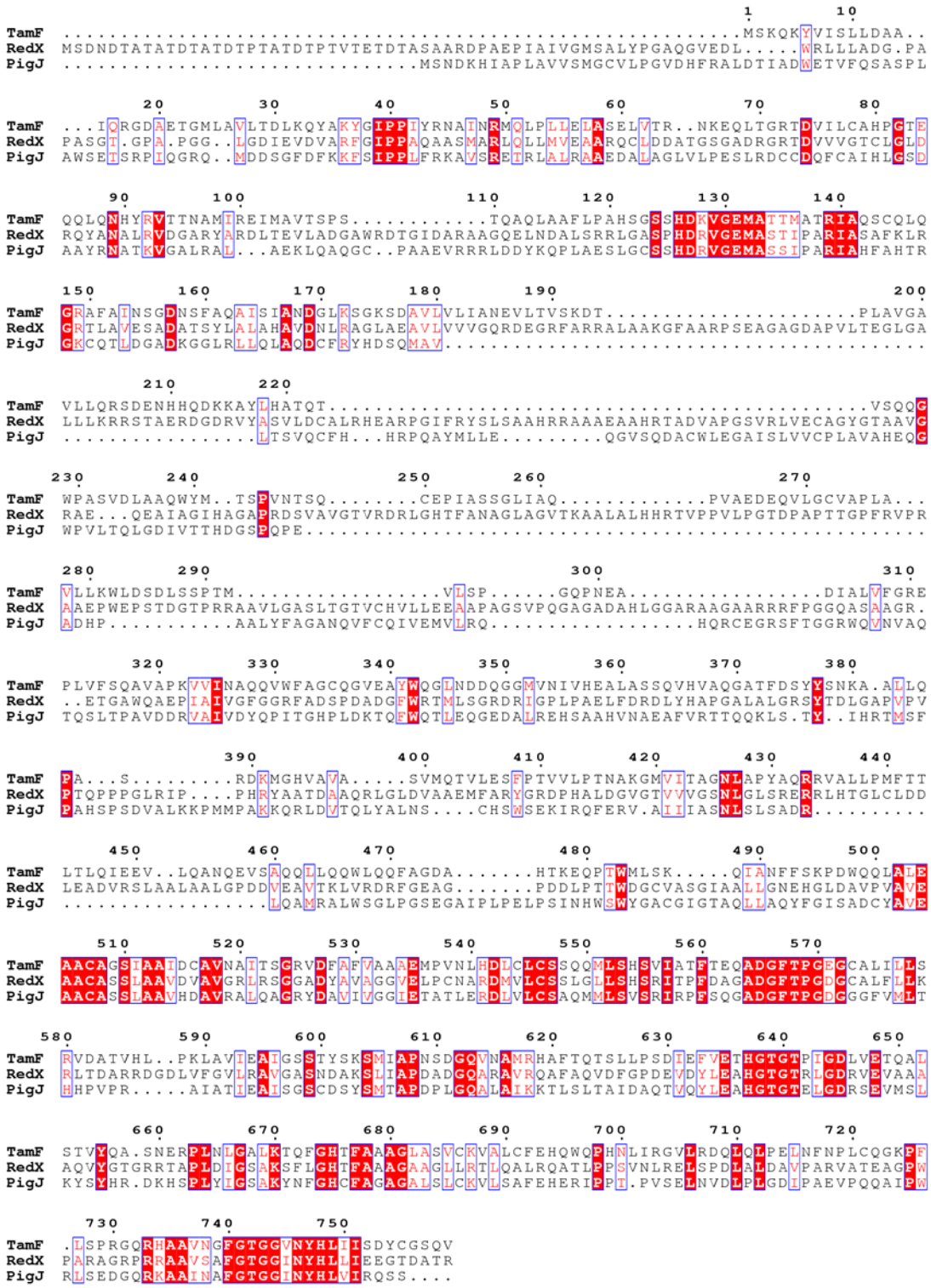
9 Appendices

Appendix 1: The *tamF* gene sequence (GenBank: EAR29364).

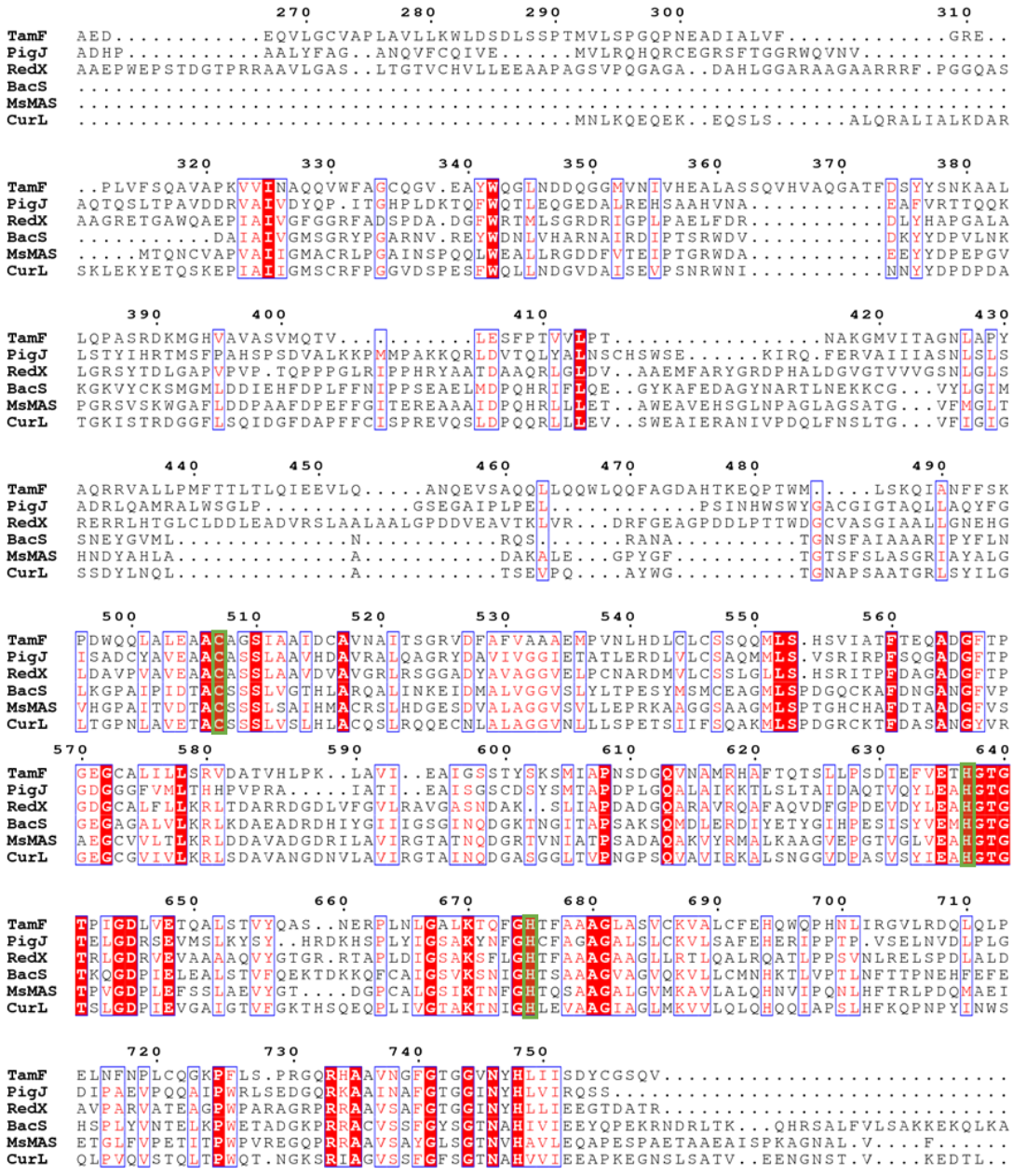
ATGAGCAAGCAGAAGTATGTGATATCGCTGCTTGATGCAGCGATACAACGAGGCGACGCCGA
GACAGGAATGCTCGCGGTGCTCACAGATTTAAAACAATATGCCAAATATGGCATCCCCCCTA
TTTACCGTAATGCAATAAATCGCATGCAATTGCCATTACTTGAGTTGGCAAGCGAGCTAGTA
ACGCGAAATAAAGAGCAGCTTACGGGTCGCACCGATGTGATTTTATGTGCTCATCCGGGCAC
CGAGCAGCAACTGCAAAATCACTACCGTGTTACCACTAATGCAATGATCCGCGAGATTATGG
CTGTTACATCGCCGAGTACGCAGGCGCAGTTAGCTGCATTTTTGCCAGCACACAGTGGCAGT
AGTCACGATAAAGTAGGTGAAATGGCAACTACTATGGCTACTCGTATTGCACAGAGCTGCCA
ATTACAGGGGCGTGCTTTTGCGATAAACAGTGGTGATAACAGTTTTGCTCAAGCTATTTCTA
TTGCCAATGATGGTTTGAAAAGCGGCAAAAGTGATGCTGTGTTGGTGCTTATTGCCAATGAA
GTGTTAACCGTTAGCAAAGACACTCCCTTGGCTGTGGGAGCTGTATTATTGCAACGAAGTGA
CGAAAATCATCACCAAGATAAAAAAGCTTATTTGCATGCTACGCAAACTGTTTCACAACAAG
GATGGCCAGCGTCTGTCGATTTAGCTGCGCAGTGGTATATGACCTCACCCGTTAATACTTCG
CAATGTGAGCCCATTGCCAGTAGCGGGCTAATTGCTCAGCCTGTCGCTGAAGATGAACAGGT
TTTAGGCTGTGTTGCACCACTTGCAGTATTGCTAAAGTGGTTAGACTCCGACTTGTCGTCGC
CAACCATGGTCTTGTCACCTGGACAGCCAAATGAAGCGGATATTGCTTTGGTCTTTGGTCGT
GAGCCATTGGTTTTCAGTCAAGCCGTTGCACCTAAAGTGGTTATCAATGCACAGCAAGTATG
GTTTGCAGGCTGCCAAGGTGTTGAAGCTTATTGGCAAGGCTTAAATGATGATCAAGGCGGTA
TGGTGAATATTGTTCATGAGGCATTAGCAAGTAGCCAAGTCCATGTTGCACAAGGGGCGACT
TTTGACAGTTATTACAGCAACAAGGCTGCACTGCTTCAACCCGCCAGTCGCGATAAAATGGG
CCATGTTGCTGTTGCGAGTGTCATGCAAACTGTGCTTGAAAGCTTTCCTACCGTTGTTTTAC
CCACGAATGCAAAAGGGATGGTGATCACCGCCGGAAATTTGGCTCCTTATGCGCAGCGACGA
GTAGCACTCTTGCCGATGTTTACTACATTAACGCTGCAGATAGAGGAAGTATTACAAGCTAA
TCAAGAAGTGTCGGCGCAGCAGCTGCTACAACAGTGGTTGCAGCAATTTGCTGGCGATGCTC
ATACCAAAGAGCAGCCGACTTGGATGTTAAGTAAACAAATTGCTAACTTTTTCTCAAAGCCT
GATTGGCAGCAATTGGCACTCGAAGCGGCATGTGCAGGCTCAATCGCGGCGATTGATTGTGC
GGTGAATGCCATCACTAGCGGCAGGGTCGATTTTGCCTTTGTTGCAGCAGCTGAAATGCCCG
TTAATTTACACGACTTATGCCTGTGTTCTAGCCAACAAATGTTATCTCATAGTGTGATTGCT
ACTTTCACCGAGCAGGCCGATGGCTTTACGCCAGGTGAAGGATGCGCCTTGATTTTACTCAG
CCGCGTTGATGCGACAGTGCATTTGCCAAAATTGGCAGTGATAGAGGCGATTGGTTCAAGCA
CCTACAGTAAATCTATGATAGCGCCAAACAGTGATGGTCAAGTAAACGCTATGCGCCATGCC
TTTACACAAACATCTTTGTTACCAAGCGATATTGAGTTTGTGGAGACTCATGGTACAGGTAC

GCCTATTGGTGACCTTGTGGAGACGCAAGCGCTTAGCACTGTGTATCAAGCAAGCAATGAAC
GGCCACTAAATCTTGGAGCCCTCAAAACCCAATTTGGCCACACTTTTGCCGCAGCAGGTTTG
GCAAGTGTGTGTAAAGTTGCGCTCTGTTTTGAGCATCAATGGCAACCGCACAATTTGATCCG
TGGCGTATTACGTGACCAATTGCAGTTGCCCGAACTTAATTTTAACCCTTTATGCCAAGGCA
AACCCTTTTTATCGCCTCGCGGACAACGTCATGCGGCGGTCAATGGCTTTGGAACCGGTGGC
GTTAACTATCACTTGATCATCAGTGACTATTGCGGATC

Appendix 2: Alignment of the TamF protein sequence with RedX (UniProtKB: O54142) and PigJ (UniProtKB: Q5W262), homologues from prodiginine biosynthesis showing that that the majority of the homology is in the C-terminal ketosynthase domain.



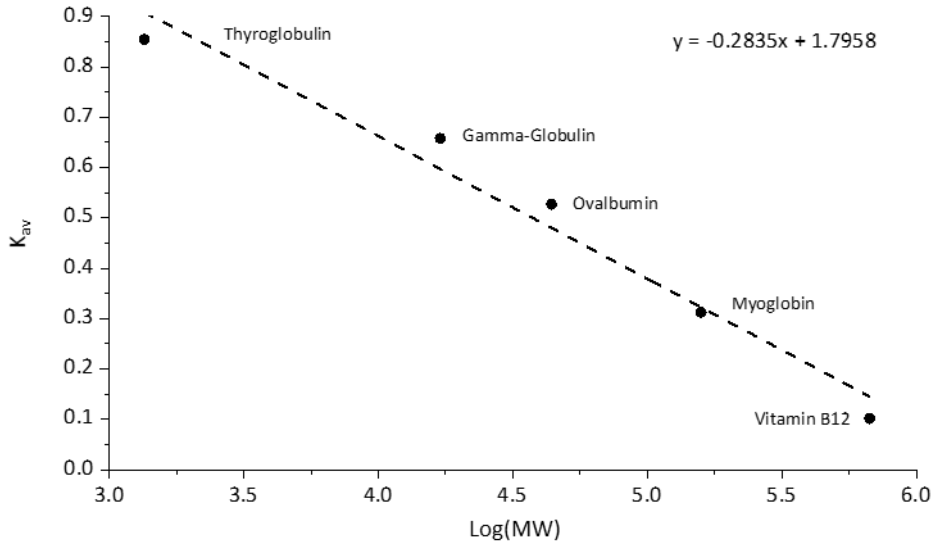
Appendix 3: Alignment of the TamF protein sequence with RedX (UniProtKB: O54142), PigJ (UniProtKB: Q5W262), *Mycobacterium smegmatis* mycocerosic acid synthase (MAS), 5BP1; *Moorea producens* CurL, 4MZ0 and *Bacillus subtilis* BacS, 5ENY. Conserved KS catalytic residues (Cys507, His637 and His675 in TamF) are highlighted in green.



Appendix 4: TamF recombinant protein sequence from pET22b expression construct including C-terminal 6xHis tag (bold) and excluding N-terminal methionine which is cleaved *in vivo* during expression.

SKQKYVISLLDAAIQRGDAETGMLAVLTDLKQYAKYGIPPIYRNAINRMQLPLLELASELVT
RNKEQLTGRTDVILCAHPGTEQQLQNHYRVTTNAMIREIMAVTSPSTQAQLAAFLPAHSGSS
HDKVGEMATTMATRIAQSCQLQGRAFAINSGDNSFAQAISIANDGLKSGKSDAVLVLIANEV
LTVSKDTPLAVGAVLLQRSDENHHQDKKAYLHATQTVSQQGWPASVDLAAQWYMTSPVNTSQ
CEPIASSGLIAQPVAEDEQVLGCVAPLAVLLKWLDSDLSSPTMVLSPGQPNEADIALVFGRE
PLVFSQAVAPKVVINAQQVWFAGCQGVEAYWQGLNDDQGGMVNIVHEALASSQVHVAQGATF
DSYYSNKAALLQPASRDKMGHVAVASVMQTVLESFPTVVLPTNAKGMVITAGNLAPYAQRRV
ALLPMFTTLTLQIEEVLQANQEVSAQQLLQQWLQQFAGDAHTKEQPTWMLSKQIANFFSKPD
WQQLALEAACAGSIAAIDCAVNAITSGRVDFAFVAAAEMPVNLHDLCLCSSQQMLSHSVIAT
FTEQADGFTPGEGCALILLSRVDATVHLPKLAVIEAIGSSTYSKSMIAPNSDGQVNAMRHAF
TQTSLLPSDIEFVETHGTGTPIGDLVETQALSTVYQASNERPLNLGALKTQFGHTFAAAGLA
SVCKVALCFEHQWQPHNLIRGVLRDQLQLPELNFNPLCQGKPFLSPRGQRHAAVNGFGTGGV
NYHLIISDYCGSQV**LEHHHHHH**

Appendix 5: Calibration curve of the Superdex HiLoad 16/60 S200 gel filtration column and resulting equation for calculating approximate molecular weight (MW) of a protein.



The MW of the protein is estimated as follows:

$$\mathrm{MW} = 10^{\frac{\mathrm{K_{av}} - 1.7958}{-0.2853}}$$

$$\mathrm{K_{av}} = \frac{\mathrm{V_e} - \mathrm{V_o}}{\mathrm{V_t} - \mathrm{V_o}}$$

Where:

MW is measured in Da

V_e = Elution volume

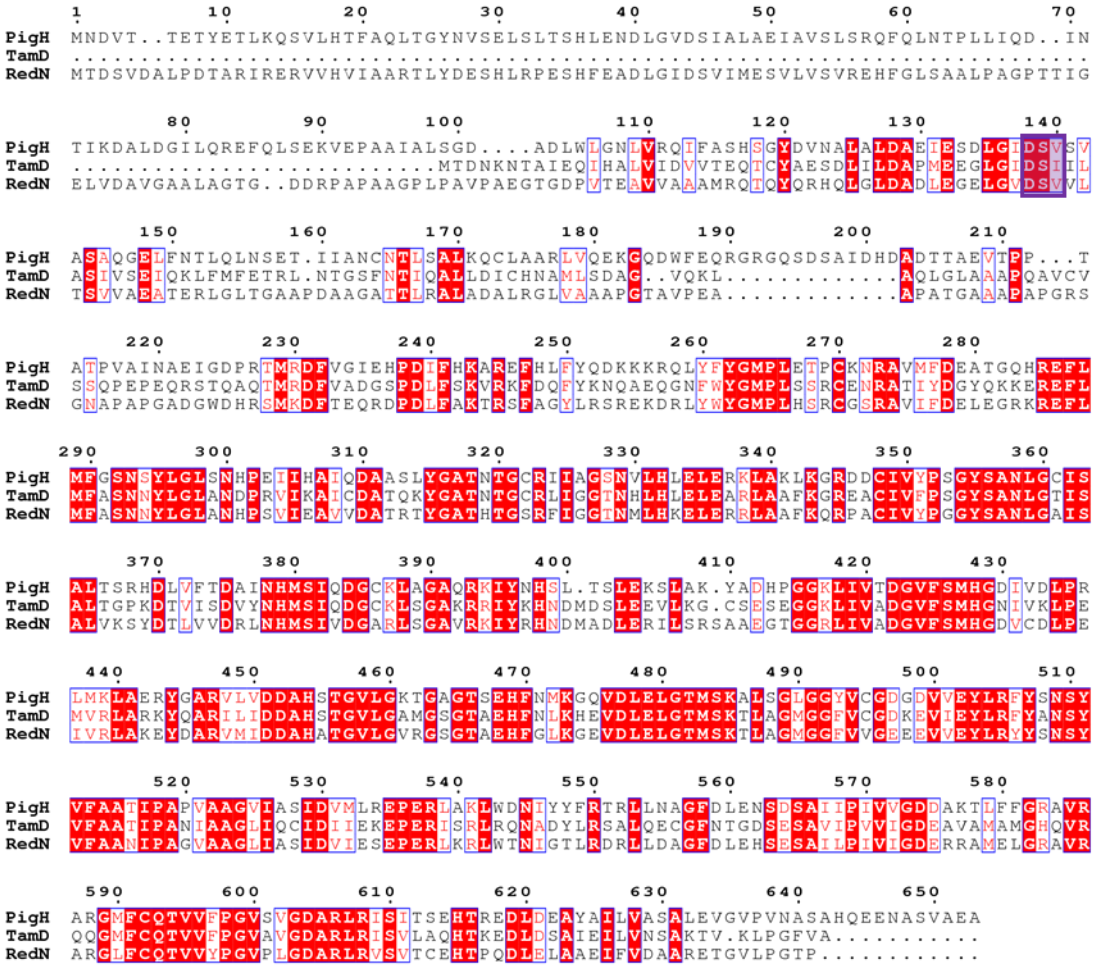
V_o = Void volume (41.9 mL)

V_t = Total bed volume (120 mL)

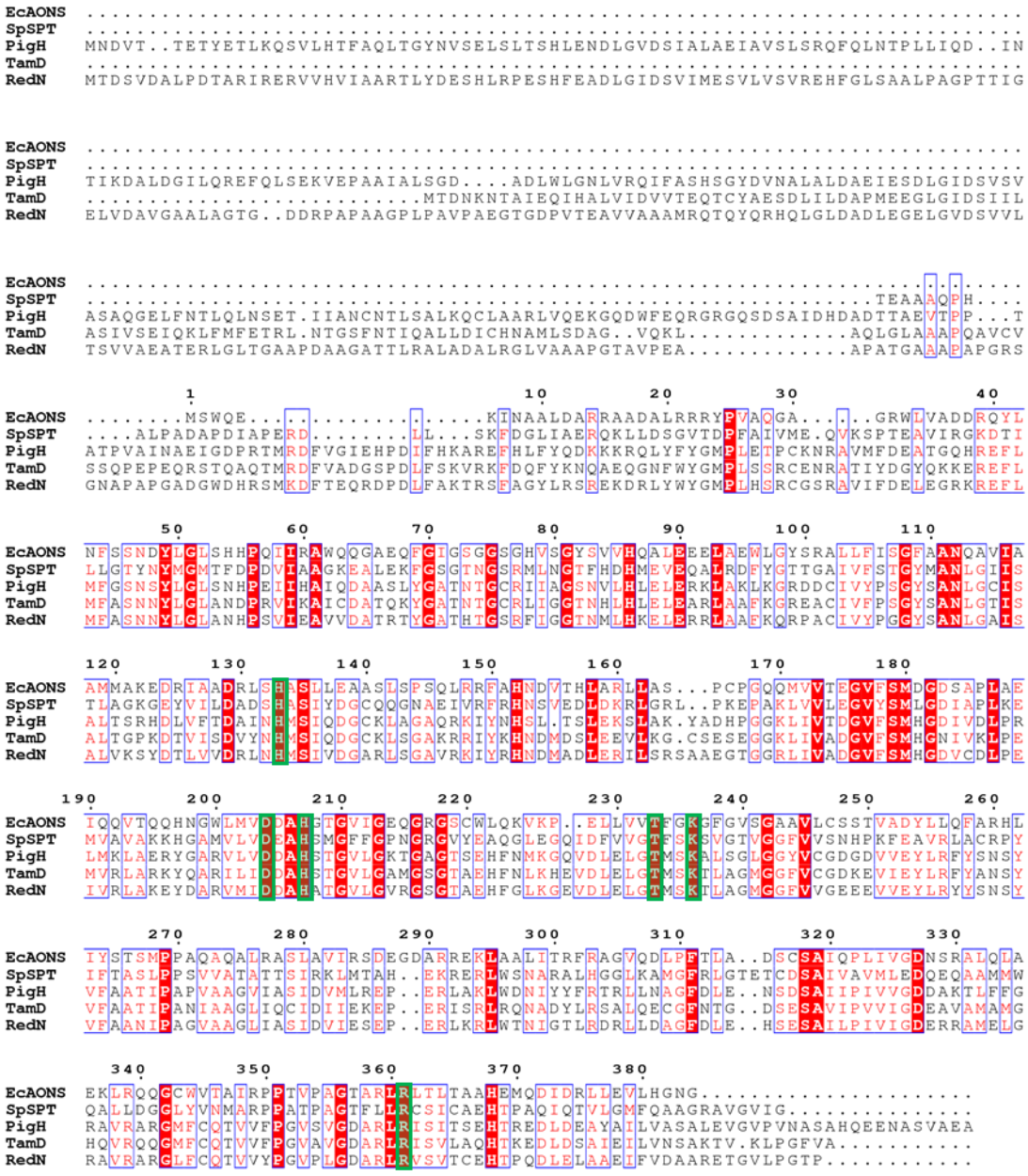
Appendix 6: The *tamD* gene sequence (GenBank: EAR29364).

ATGACAGACAATAAAAACACCGCTATTGAGCAAATACACGCCCTAGTTATCGACGTAGTAAC
AGAGCAAACGTGTTATGCCGAGTCGGATTTAATTTTAGATGCGCCGATGGAAGAAGGCTTGG
GGATAGACTCCATTATCCTTGCTTCTATCGTCAGTGAAATTCAAAAATTGTTTATGTTTGAG
ACCCGTCTCAATACTGGCAGTTTTAATACCATTCAAGCATTACTCGACATTTGTCACAATGC
GATGCTATCAGACGCAGGAGTGCAAAAACTGGCACAATTAGGACTTGCAGCAGCACCACAAG
CTGTTTGTGTAAGTTCGCAGCCAGAGCCTGAACAGCGTTCAACTCAGGCACAAACAATGCGA
GATTTTGTCGCAGATGGTAGCCCTGACTTATTTAGTAAAGTGCGTAAGTTTGACCAGTTTTA
TAAAAATCAGGCTGAGCAAGGTAACTTTTGGTACGGCATGCCACTTAGCTCCAGATGTGAAA
ATCGAGCGACTATTTATGATGGCTATCAGAAAAAAGAACGTGAATTCTTAATGTTTGCCTCG
AATAATTATTTGGGGTTAGCTAACGACCCGCGGGTCATCAAAGCAATCTGTGATGCTACGCA
AAAATACGGTGCAACAAATACAGGTTGTCGTCTGATTGGTGGCACTAATCATTTGCACCTTG
AACTGGAAGCACGTTTGGCAGCGTTTAAAGGTCGCGAAGCCTGTATTGTTTTCCCCTCTGGT
TATTCGGCTAACCTTGGTACGATTTCTGCGTTAACTGGTCCAAAAGACACTGTGATTTCAGA
TGTTTATAATCACATGAGTATTCAAGATGGTTGTAAGTTATCAGGTGCAAAACGCCGTATTT
ACAAACATAACGATATGGATTCGTTAGAAGAAGTACTAAAGGGTTGCAGTGAGTCCGAAGGT
GGTAAGCTAATTGTTGCTGATGGTGTGTTCAGCATGCACGGTAACATTGTAAAACTGCCAGA
AATGGTGCGTCTGGCCCGTAAATATCAAGCACGTATTTTGATTGATGATGCCCATTCTACTG
GGGTGTTAGGGGCAATGGGGTCAGGTACTGCTGAGCATTTCAACCTCAAACATGAAGTCGAT
CTTGAATTAGGCACCATGAGTAAAACGCTGGCTGGTATGGGGGGATTTGTCTGTGGTGACAA
AGAAGTAATTGAGTATTTACGTTTTTACGCAAATTCTTATGTGTTTGCTGCAACTATTCCTG
CAAATATCGCGGCGGGTTTAATTCAATGTATTGATATTATTGAAAAAGAGCCAGAGAGAATT
AGTCGTTTACGCCAAAATGCAGATTACCTGCGCAGTGCATTACAAGAATGCGGTTTTAATAC
CGGTGACAGTGAAAGTGCCGTCATTCCTGTGGTTATTGGCGATGAAGCTGTAGCAATGGCAA
TGGGCCATCAAGTTCGTCAGCAAGGCATGTTCTGTCAAACGGTAGTATTTCCGGGTGTTGCA
GTTGGTGATGCACGTTTACGGATCAGCGTTCTGGCTCAGCATACAAAAGAAGACCTCGATAG
TGCCATTGAGATTTTAGTGAACTCAGCTAAAACAGTGAAATTACCCGGTTTTGTGGCATAG

Appendix 7: Alignment of TamD protein sequence with homologues from prodiginine biosynthesis, RedN (UniProtKB: O54153) and PigH (UniProtKB: Q5W264). Alignment shows that the TamD ACP is homologous to the second RedN/PigH ACP domains. The conserved DSX motif for 4'-phosphopantetheine (4'-PP) binding is shown in purple.



Appendix 8: Alignment of TamD protein sequence with RedN (UniProtKB: O54153), PigH (UniProtKB: Q5W264), *Sphingomonas paucimobilis* SPT, 2JG2 and *Escherichia coli* AONS, 1DJE (UniProtKB: P12998). Conserved residues involved in PLP binding and catalysis in the AOS domain are highlighted in green.



Appendix 9: TamD recombinant protein sequence from pEHISTEV expression construct. Tobacco Etch Virus (TEV) protease cleavage removes the residues shown in bold including the 6xHis tag.

MSYYHHHHHHDYDIPTTENLYFQGAMADNKNTAIEQIHALVIDVVTEQTCYAESDLILDAPM
EEGLGIDSIILASIVSEIQKLFMFETRLNTGSFNTIQALLDICHNAMLSDAGVQKLAQLGLA
AAPQAVCVSSQPEPEQRSTQAQTMRDFVADGSPDLFSKVRKFDQFYKNQAEQGNFWYGMPLS
SRCENRATIYDGYQKKEREFLMFASNNYLGLANDPRVIKAICDATQKYGATNTGCRLIGGTN
HLHLELEARLAAFKGREACIVFPSGYSANLGTISALTGPKDTVISDVYNHMSIQDGCKLSGA
KRRIYKHNDMDSLEEVLKGCSESEGGKLIVADGVFSMHGNIVKLPEMVRLARKYQARILIDD
AHSTGVLGAMGSGTAEHFNLKHEVDLELGTMSKTLAGMGGFVCGDKEVIEYLRFYANSYVFA
ATIPANIAAGLIQCIDIIEKEPERISRLRQNADYLRSALQECGFNTGDSESAVIPVVIGDEA
VAMAMGHQVRQQGMFCQTVVFPGVAVGDARLRISVLAQHTKEDLDSAIEILVNSAKTVKLPG
FVA

Appendix 10: TamD ACP recombinant protein sequences from pEHISTEV expression constructs. Tobacco Etch Virus (TEV) protease cleavage removes the residues shown in bold including the 6xHis tag. TamD ACP_{98} was the construct utilized in further experiments.

>TamD ACP_{95}

MSYYHHHHHHDYDIPTTENLYFQGAMADNKNTAIEQIHALVIDVVTEQTCYAESDLILDAPM
EEGLGIDSIILASIVSEIQKLFMFETRLNTGSFNTIQALLDICHNAMLSDAGVQKLAQ

>TamD ACP_{98}

MSYYHHHHHHDYDIPTTENLYFQGAMADNKNTAIEQIHALVIDVVTEQTCYAESDLILDAPM
EEGLGIDSIILASIVSEIQKLFMFETRLNTGSFNTIQALLDICHNAMLSDAGVQKLAQLGL

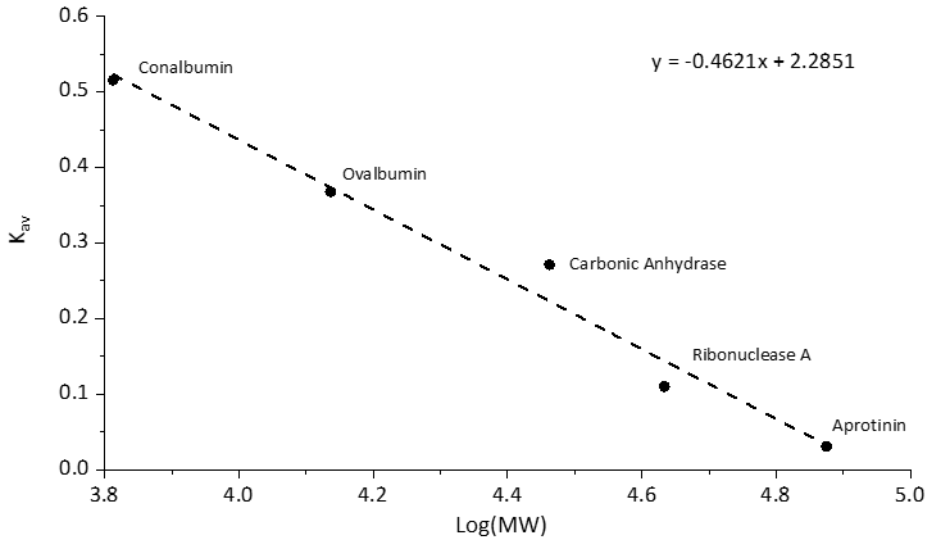
>TamD ACP_{101}

MSYYHHHHHHDYDIPTTENLYFQGAMADNKNTAIEQIHALVIDVVTEQTCYAESDLILDAPM
EEGLGIDSIILASIVSEIQKLFMFETRLNTGSFNTIQALLDICHNAMLSDAGVQKLAQLGLA
AA

>TamD ACP_{104}

MSYYHHHHHHDYDIPTTENLYFQGAMADNKNTAIEQIHALVIDVVTEQTCYAESDLILDAPM
EEGLGIDSIILASIVSEIQKLFMFETRLNTGSFNTIQALLDICHNAMLSDAGVQKLAQLGLA
AAPQA

Appendix 11: Calibration curve of the Superdex HiLoad 16/60 S75 gel filtration column and resulting equation for calculating approximate molecular weight (MW) of a protein.



The MW of the protein is estimated as follows:

$$\mathrm{MW} = 10^{\frac{\mathrm{K_{av}} - 1.7958}{-0.2853}}$$

$$\mathrm{K_{av}} = \frac{\mathrm{V_e} - \mathrm{V_o}}{\mathrm{V_t} - \mathrm{V_o}}$$

Where:

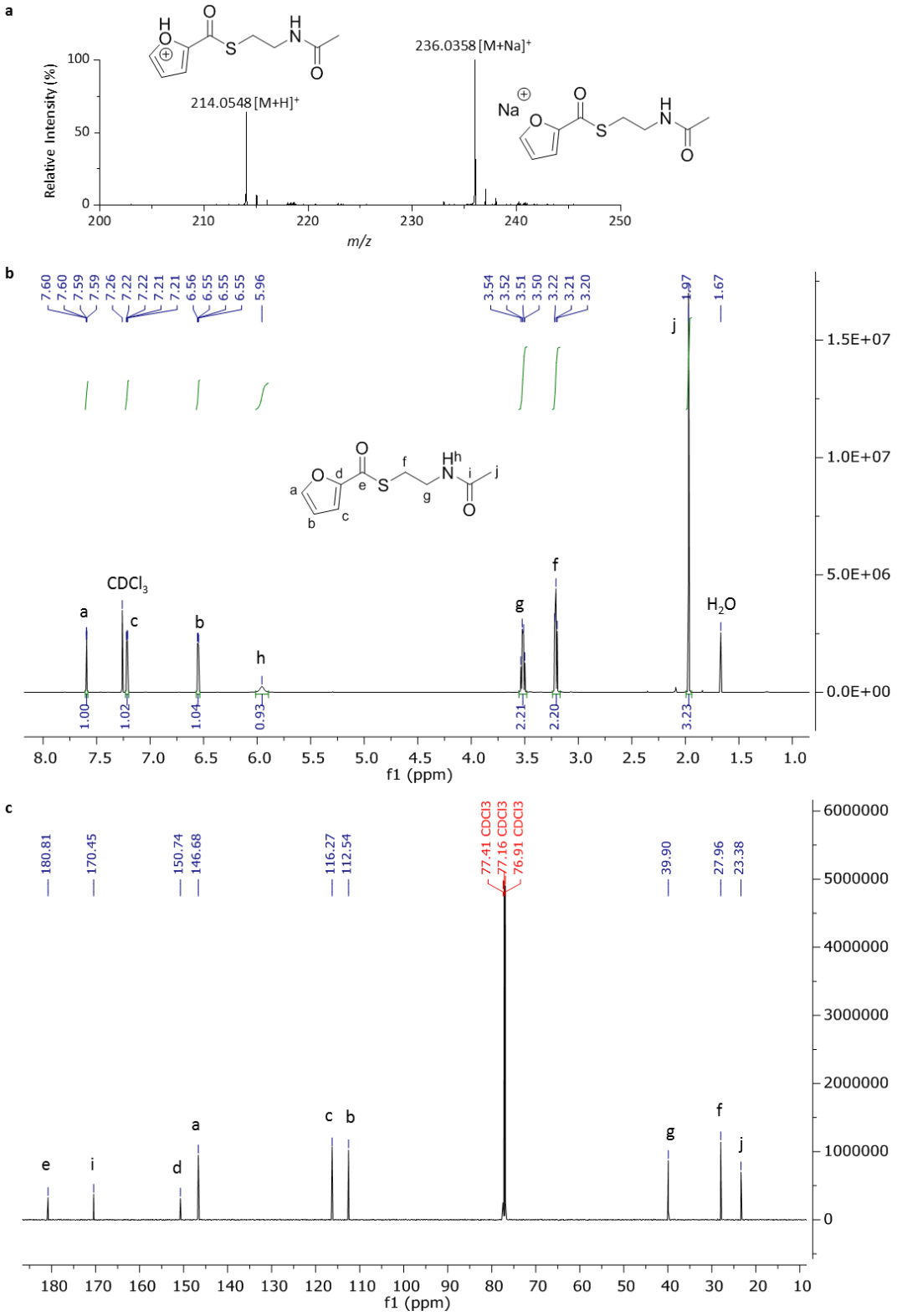
MW is measured in Da

V_e = Elution volume

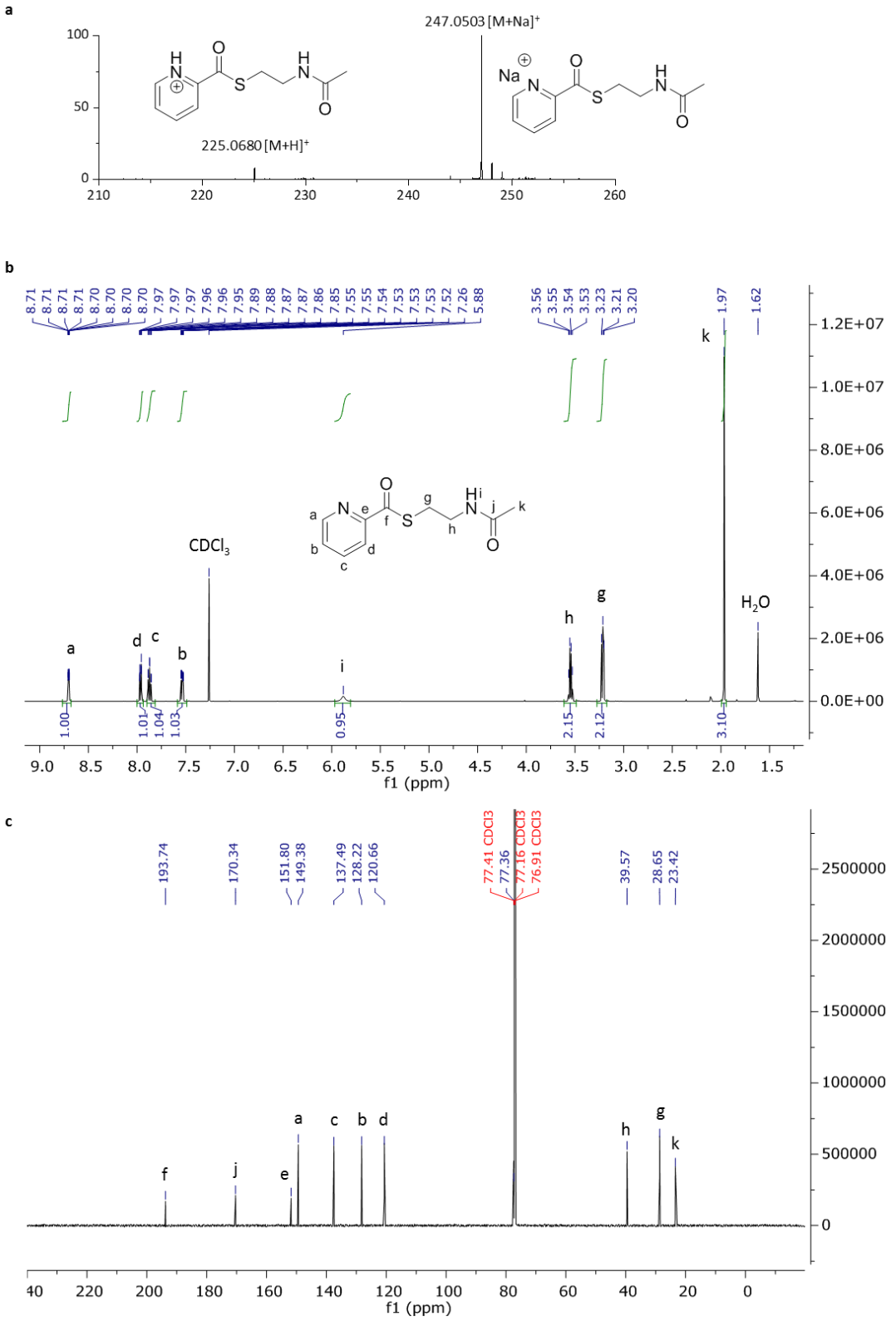
V_o = Void volume (41.9 mL)

V_t = Total bed volume (120 mL)

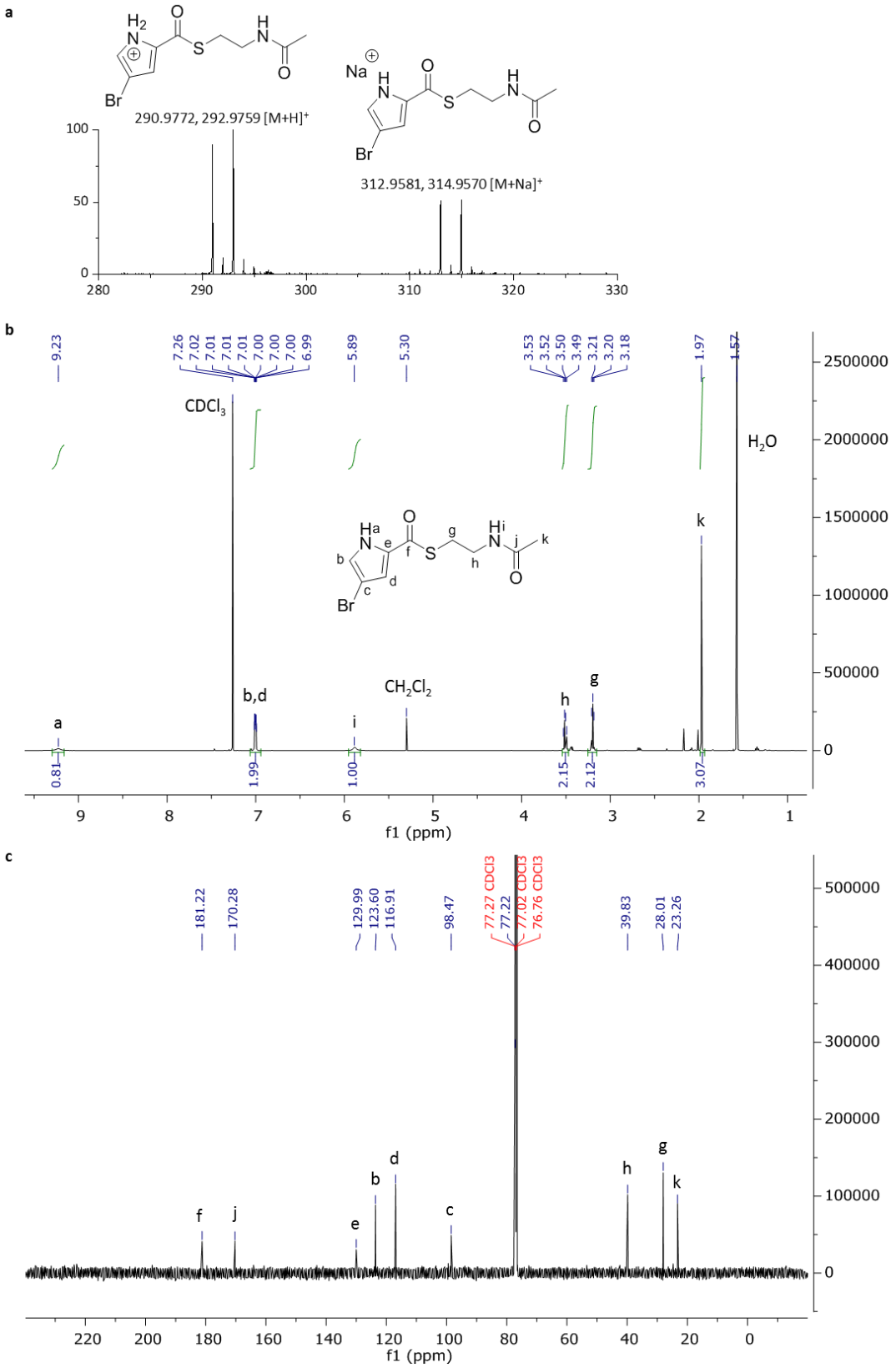
Appendix 12: Characterization of furan SNAC (**a**) high resolution MS, (**b**) ^{1}H NMR and (**c**) ^{13}C NMR



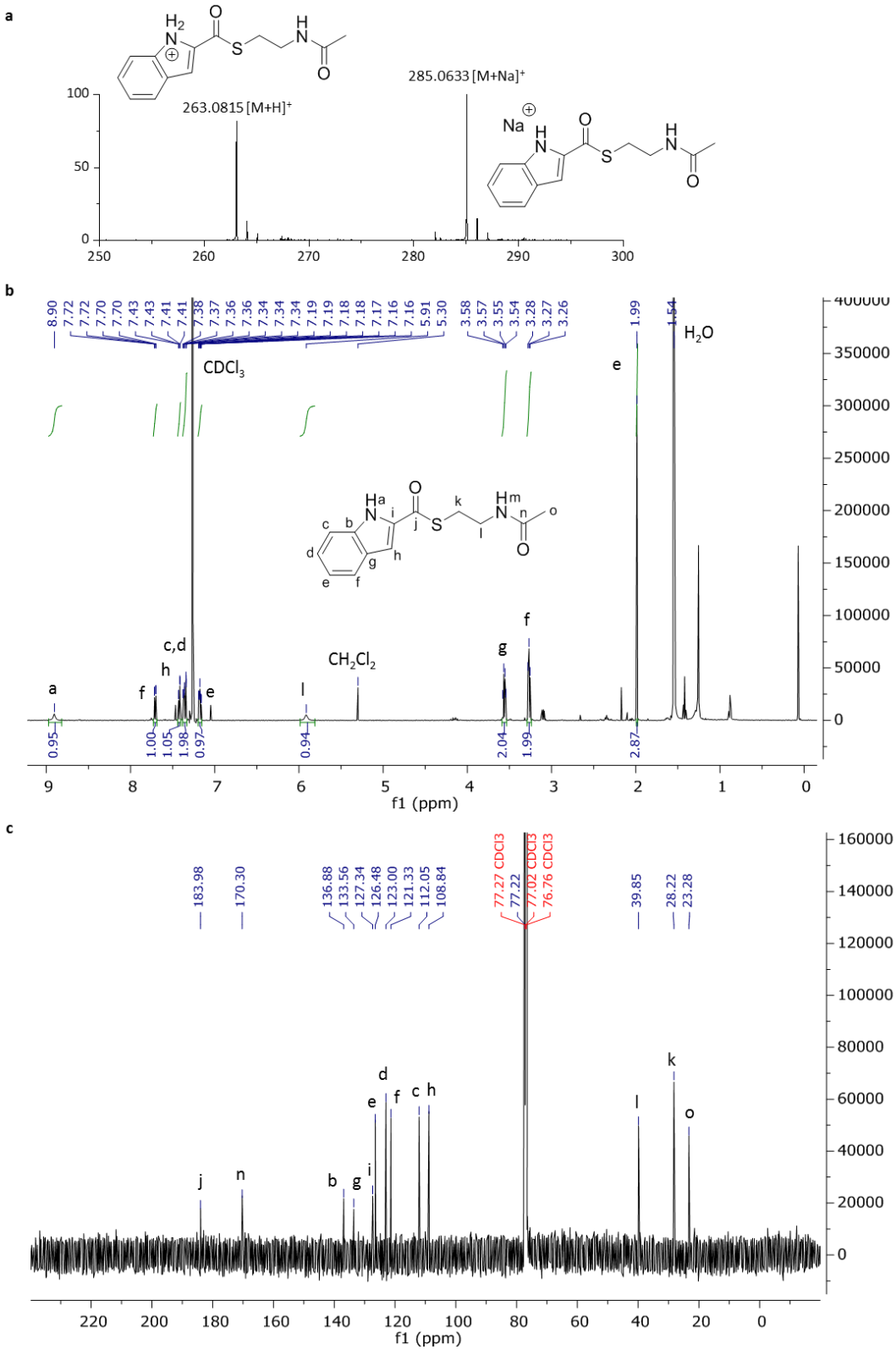
Appendix 13: Characterization of pyridine SNAC (**a**) high resolution MS, (**b**) ^{1}H NMR and (**c**) ^{13}C NMR



Appendix 14: Characterization of 4-bromopyrrole SNAC (**a**) HRMS, (**b**) ^{1}H NMR and (**c**) ^{13}C NMR



Appendix 15: Characterization of indole SNAC (**a**) HRMS, (**b**) ^{1}H NMR and (**c**) ^{13}C NMR.



Appendix 16: TamD $AOS_{\Delta 95}$ recombinant protein sequence from expression in pET22b including a C-terminal 6xHis tag (bold).

MLGLAAAPQAVCVSSQPEPEQRSTQAQTMRDFVADGSPDLFSKVRKFDQFYKNQAEQGNFWY
GMPLSSRCENRATIYDGYQKKEREFLMFASNNYLGLANDPRVIKAICDATQKYGATNTGCRL
IGGTNHLHLELEARLAAFKGREACIVFPSGYSANLGTISALTGPKDTVISDVYNHMSIQDGC
KLSGAKRRIYKHNDMDSLEEVLKGCSESEGGKLIVADGVFSMHGNIVKLPEMVRLARKYQAR
ILIDDAHSTGVLGAMGSGTAEHFNLKHEVDLELGTMSKTLAGMGGFVCGDKEVIEYLRFYAN
SYVFAATIPANIAAGLIQCIDIIEKEPERISRLRQNADYLRSALQECGFNTGDSESAVIPVV
IGDEAVAMAMGHQVRQQGMFCQTVVFPGVAVGDARLRISVLAQHTKEDLDSAIEILVNSAKT
VKLPGFVA**LEHHHHHH**

Appendix 17: TamD AOS recombinant protein sequences from expression in pET22b including a C-terminal 6xHis tag (bold) with different linker region truncations. None of these constructs without the linker region produced soluble protein.

>TamD $AOS_{\Delta 122}$

TMRDFVADGSPDLFSKVRKFDQFYKNQAEQGNFWYGMPLSSRCENRATIYDGYQKKEREFLM
FASNNYLGLANDPRVIKAICDATQKYGATNTGCRLIGGTNHLHLELEARLAAFKGREACIVF
PSGYSANLGTISALTGPKDTVISDVYNHMSIQDGCKLSGAKRRIYKHNDMDSLEEVLKGCSE
SEGGKLIVADGVFSMHGNIVKLPEMVRLARKYQARILIDDAHSTGVLGAMGSGTAEHFNLKH
EVDLELGTMSKTLAGMGGFVCGDKEVIEYLRFYANSYVFAATIPANIAAGLIQCIDIIEKEP
ERISRLRQNADYLRSALQECGFNTGDSESAVIPVVIGDEAVAMAMGHQVRQQGMFCQTVVFP
GVAVGDARLRISVLAQHTKEDLDSAIEILVNSAKTVKLPGFVA**LEHHHHHH**

>TamD $AOS_{\Delta 125}$

DFVADGSPDLFSKVRKFDQFYKNQAEQGNFWYGMPLSSRCENRATIYDGYQKKEREFLMFAS
NNYLGLANDPRVIKAICDATQKYGATNTGCRLIGGTNHLHLELEARLAAFKGREACIVFPSG
YSANLGTISALTGPKDTVISDVYNHMSIQDGCKLSGAKRRIYKHNDMDSLEEVLKGCSESEG
GKLIVADGVFSMHGNIVKLPEMVRLARKYQARILIDDAHSTGVLGAMGSGTAEHFNLKHEVD
LELGTMSKTLAGMGGFVCGDKEVIEYLRFYANSYVFAATIPANIAAGLIQCIDIIEKEPERI
SRLRQNADYLRSALQECGFNTGDSESAVIPVVIGDEAVAMAMGHQVRQQGMFCQTVVFPGVA
VGDARLRISVLAQHTKEDLDSAIEILVNSAKTVKLPGFVA**LEHHHHHH**

>TamD $AOS_{\Delta 128}$

ADGSPDLFSKVRKFDQFYKNQAEQGNFWYGMPLSSRCENRATIYDGYQKKEREFLMFASNNY
LGLANDPRVIKAICDATQKYGATNTGCRLIGGTNHLHLELEARLAAFKGREACIVFPSGYSA
NLGTISALTGPKDTVISDVYNHMSIQDGCKLSGAKRRIYKHNDMDSLEEVLKGCSESEGGKL
IVADGVFSMHGNIVKLPEMVRLARKYQARILIDDAHSTGVLGAMGSGTAEHFNLKHEVDLEL
GTMSKTLAGMGGFVCGDKEVIEYLRFYANSYVFAATIPANIAAGLIQCIDIIEKEPERISRL
RQNADYLRSALQECGFNTGDSESAVIPVVIGDEAVAMAMGHQVRQQGMFCQTVVFPGVAVGD
ARLRISVLAQHTKEDLDSAIEILVNSAKTVKLPGFVA**LEHHHHHH**

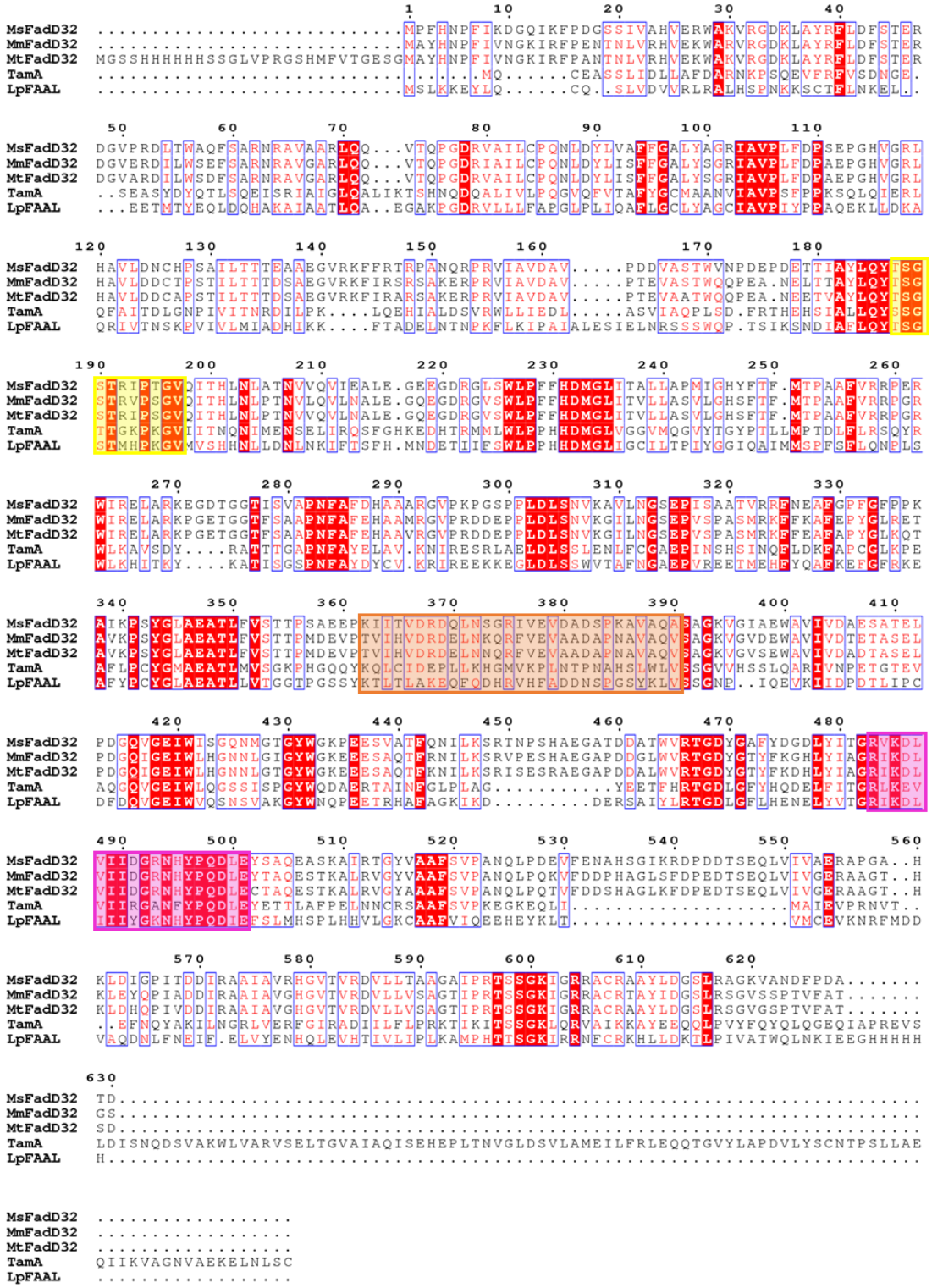
>TamD $AOS_{\Delta 131}$

SPDLFSKVRKFDQFYKNQAEQGNFWYGMPLSSRCENRATIYDGYQKKEREFLMFASNNYLGL
ANDPRVIKAICDATQKYGATNTGCRLIGGTNHLHLELEARLAAFKGREACIVFPSGYSANLG
TISALTGPKDTVISDVYNHMSIQDGCKLSGAKRRIYKHNDMDSLEEVLKGCSESEGGKLIVA
DGVFSMHGNIVKLPEMVRLARKYQARILIDDAHSTGVLGAMGSGTAEHFNLKHEVDLELGTM
SKTLAGMGGFVCGDKEVIEYLRFYANSYVFAATIPANIAAGLIQCIDIIEKEPERISRLRQN
ADYLRSALQECGFNTGDSESAVIPVVIGDEAVAMAMGHQVRQQGMFCQTVVFPGVAVGDARL
RISVLAQHTKEDLDSAIEILVNSAKTVKLPGFVA**LEHHHHHH**

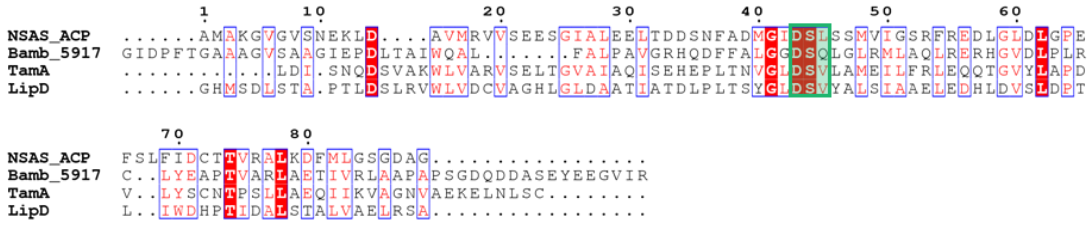
Appendix 18: The *tamA* gene sequence (GenBank: EAR29369).

ATGCAATGTGAAGCAAGTAGTTTGATTGACTTACTTGCTTTTGATGCTCGAAACAAGCCGAG
CCAAGAAGTATTTAGATTTGTTAGTGATAATGGTGAGTCAGAGGCTTCGTATGATTATCAAA
CTCTTAGCCAAGAAATAAGCCGAATTGCGATTGGGCTACAAGCATTAATCAAGACAAGTCAT
AATCAAGATCAGGCGTTGATTGTGCTCCCGCAAGGGGTGCAATTTGTGACCGCTTTTTATGG
CTGTATGGCAGCGAATGTGATTGCTGTGCCATCTTTCCCCCCCAAAAGTCAGTTACAAATTG
AACGCTTACAGTTTGCCATTACTGATTTAGGTAACCCAATAGTTATCACCAATCGCGATATT
TTACCCAAATTACAAGAACATATTGCACTTGACTCGGTACGCTGGTTATTAATTGAAGACTT
AGCAAGTGTCATTGCACAACCGCTTAGTGATTTTAGAACCCATGAACACTCAATTGCATTGC
TGCAATATAGCTCTGGTACTACAGGTAAGCCTAAAGGGGTCATTATCACCAACCAAAACATT
ATGGAAAACTCGGAGCTGATCCGACAAAGTTTTGGCCATAAAGAAGACCATACCAGAATGAT
GCTGTGGTTGCCTCCGCATCATGATATGGGGTTAGTGGGCGGGGTTATGCAAGGTGTTTATA
CTGGTTATCCGACATTATTAATGCCGACCGATTTGTTCTTGCGTAGTCAATATCGCTGGTTA
AAAGCGGTCAGTGATTATCGTGCAACAACGACAGGTGCGCCTAACTTCGCTTATGAATTAGC
AGTTAAGAATATCCGAGAATCGCGCCTAGCGGAATTAGACTTGTCCTCGTTAGAAAATCTGT
TTTGTGGTGCAGAGCCCATTAATTCGCATTCTATTAATCAATTTCTTGATAAGTTTGCGCCA
TGTGGACTGAAGCCTGAAGCATTTTTACCTTGTTATGGTATGGCAGAGGCCACGTTAATGGT
TAGTGGTAAGCCTCATGGGCAACAGTACAAGCAATTATGTATCGATGAGCCGTTACTTAAAC
ATGGCATGGTCAAGCCGTTAAATACTCCAAACGCCCACAGCCTTTGGTTAGTCAGTAGTGGA
GTGGTGCATAGCTCGTTACAAGCACGTATCGTCAATCCAGAAACAGGTACTGAAGTTGCGCA
AGGTCAAGTTGGTGAAATATGGTTGCAAGGCAGTAGTATTAGCCCTGGCTATTGGCAAGATG
CAGAACGTACCGCAATCAATTTTGGTTTACCGCTCGCTGGATACGAAGAGACTTTCCACCGC
ACTGGCGACCTTGGTTTTTATCATCAAGATGAACTTTTTATTACTGGCCGCTTAAAAGAAGT
CGTGATCATTCGTGGGGCGAATTTCTATCCGCAAGATCTTGAATACGAAACAACGCTGGCTT
TTCCTGAATTAAATAATTGCCGCAGTGCTGCGTTTTCTGTGCCTAAAGAGGGCAAAGAACAA
TTGATTATGGCGATTGAAGTGCCGCGCAATGTGACTGAATTTAATCAGTACGCCAAAATTTT
AAATGGTCGGTTAGTAGAGCGCTTTGGTATCCGTGCCGACATTATTCTGTTTTTACCTCGTA
AAACGATAAAAATTACCTCTTCTGGCAAGTTGCAGCGTGTTGCAATTAAAAAAGCCTATGAA
GAACAACAACTGCCGGTGTATTTCCAATATCAACTGCAAGGCGAGCAAATTGCGCCACGTGA
AGTGTCGCTTGATATCAGCAATCAAGACAGTGTTGCTAAGTGGTTAGTTGCCCGAGTAAGTG
AATTAACAGGGGTGGCAATCGCTCAAATTTCTGAGCATGAGCCATTAACCAATGTTGGTTTG
GATTCAGTATTAGCGATGGAAATTCTGTTTCGGCTTGAGCAGCAAACAGGAGTTTATTTAGC
GCCTGATGTGCTTTACAGCTGCAATACGCCGAGCTTATTGGCAGAGCAAATCATCAAAGTCG
CAGGAAATGTAGCAGAAAAGGAGCTGAATTTATCATGTTAG

Appendix 19: Alignment of TamA protein sequence with *Mycobacterium smegmatis* FadD32, 5D6J; *Mycobacterium tuberculosis* FadD32, 5HM3; *Mycobacterium marinum* FadD32, 5EY9 and *Legionella pneumophila* FAAL, 3KXW). The conserved ATP binding P-loop is highlighted in yellow, the FAAL insertion loop in orange and the hinge region in pink.



Appendix 20: Alignment of the predicted TamA ACP protein sequence with *Actinoplanes friulienses* LipD, 2N98; *Burkholderia ambifaria* Bamb_5917, 5MTI and the *Aspergillus parasiticus* norsolorinic acid synthase (NSAS), 2KR5. The conserved DSX motif is highlighted in green.



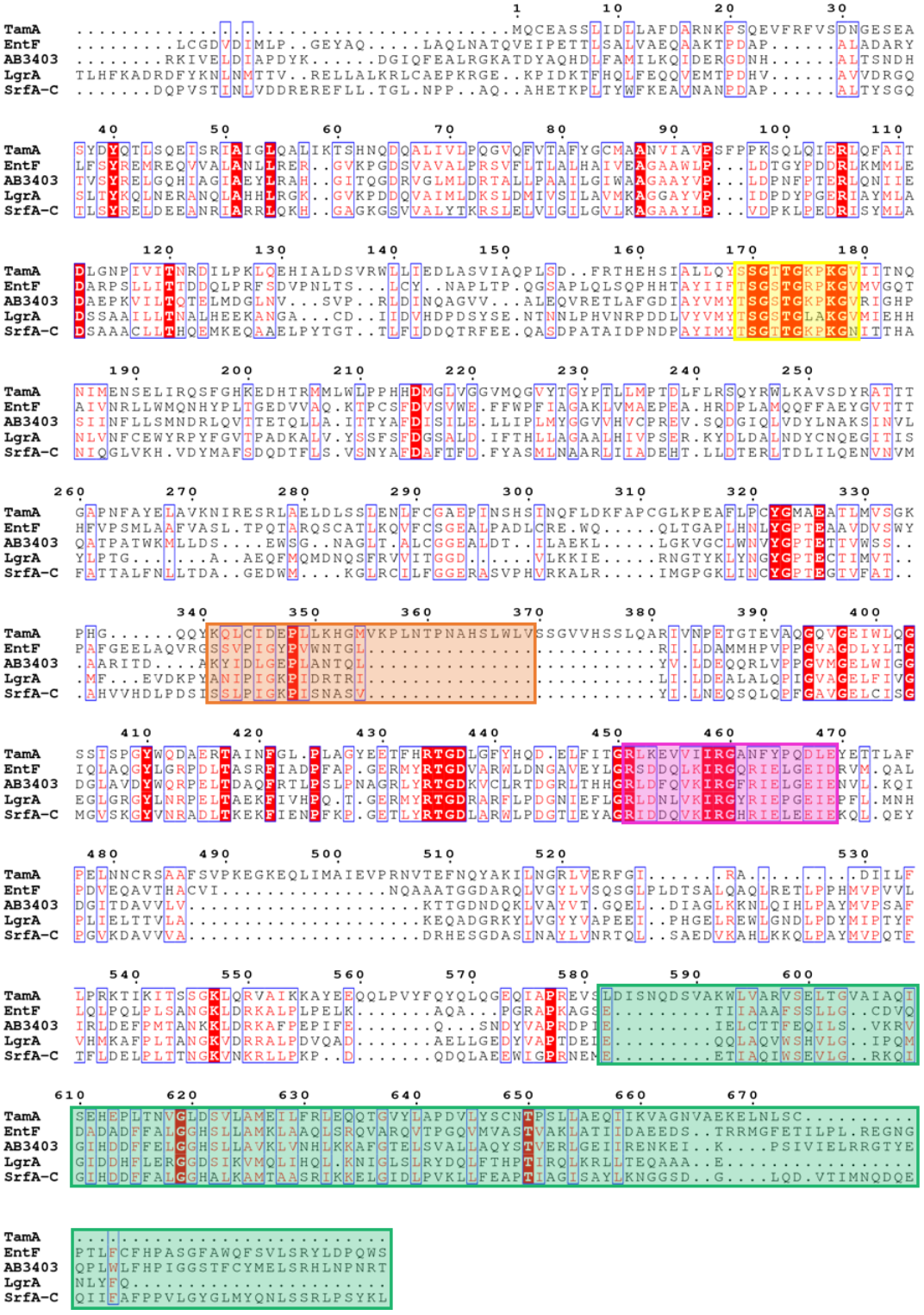
Appendix 21: TamA recombinant protein sequence from pEHISTEV expression construct. Tobacco Etch Virus (TEV) protease cleavage removes the residues shown in bold including the 6xHis tag.

MSYYHHHHHHDYDIPTTENLYFQGAMECEASSLIDLLAFDARNKPSQEVFRFVSDNGESEAS
YDYQTLSQEISRIAIGLQALIKTSHNQDQALIVLPQGVQFVTAFYGCMAANVIAVPSFPPKS
QLQIERLQFAITDLGNPIVITNRDILPKLQEHIALDSVRWLLIEDLASVIAQPLSDFRTHEH
SIALLQYSSGTTGKPKGVIITNQNIMENSELIRQSFGHKEDHTRMMLWLPPHHDMGLVGGVM
QGVYTGYPTLLMPTDLFLRSQYRWLKAVSDYRATTTGAPNFAYELAVKNIRESRLAELDLSS
LENLFCGAEPINSHSINQFLDKFAPCGLKPEAFLPCYGMAEATLMVSGKPHGQQYKQLCIDE
PLLKHGMVKPLNTPNAHSLWLVSSGVVHSSLQARIVNPETGTEVAQGQVGEIWLQGSSISPG
YWQDAERTAINFGLPLAGYEETFHRTGDLGFYHQDELFITGRLKEVVIIRGANFYPQDLEYE
TTLAFPELNNCRSAAFSVPKEGKEQLIMAIEVPRNVTEFNQYAKILNGRLVERFGIRADIIL
FLPRKTIKITSSGKLQRVAIKKAYEEQQLPVYFQYQLQGEQIA

Appendix 22: TamA ACP recombinant protein sequence from pEHISTEV expression construct. Tobacco Etch Virus (TEV) protease cleavage removes the residues shown in bold including the 6xHis tag.

MSYYHHHHHHDYDIPTTENLYFQGAMAPREVSLDISNQDSVAKWLVARVSELTGVAIAQISE
HEPLTNVGLDSVLAMEILFRLEQQTGVYLAPDVLYSCNTPSLLAEQIIKVAGNVAEKELNLS
C

Appendix 23: Alignment of TamA protein sequence with *E. coli* EntF: 5T3D, 5JA2; *Acinetobacter baumannii* AB3403, 4ZXI; *Brevibacillus parabrevis* LgrA, 5ES8 and *Geobacillus sp.* strain Y4.1MC1 SrfA-C, 5U89, 2VSQ. The conserved ATP binding P-loop is highlighted in yellow, the FAAL insertion loop in orange, the hinge region in pink and the carrier proteins in green.



Appendix 24: The *tamT* gene sequence (GenBank: EAR29351) and the resulting TamT protein sequence (UniProtKB: A4C5U7).

>*tamT*

ATGCATGACGATAATACTTTTTATGCTGCGGCGGCCAAGTTAGAACAACAATGTGGCGATCC
GAGCGATGAACAGCAATTTATAAATTTTAAAACCTCGAGTTTATTAGATGAAACTCAGCAAT
TTCCAACAGCCTACATTGAATATCTGCAACAGCTTGGCTATTTCGAATATTTTGTACCCAGC
GAGCAAGGTGGACAGCTGAAGTCTTTCGCCCAGTTACTTTTAATTTGTAGAGCGTTATCGCG
CCGTGACCTTAATGCTGCCATTGCCACAGGACAGACAATGTTAGGCGCTTTGCCAGTGTGGT
TATGCGGTTCTTCGGCGCAAAAAAATCAGTTAGCCAGCCATATTTTGCAAGGGCATTTAGGC
TGCTTGGCACTGACAGAAAAACGCCATGGTAGCAACCTACTCGCCAGCGAAGTAAGTGCAAC
TGCAACAGCTACGGGTTATGAGTTAAGTGGCGAAAAATGGCTAATTAATAATGCAACTAAAG
GTCACACTATGACCTTATTAGCGCGTAAACAAGAAACCAGTGGGCGCTCTGAACTGGCATTA
TTATTTATCGATAAAACAAACTTAACCGAATTTGAAAACCTCGATAAAATTAATACCCATGG
TATTCGCTGTGCTGATATTAGTGGGATCCGTTTTAATAATACCGAAGTGGCGAGCGATGTGG
TGATCAAAACCGCAGAACCTGGGATCTATGGCGTAATGAAAGCGCTGCAAATTTCACGGATC
CTATGTGCTGGTTTTTCATTAGGGGCGCTTGATACATGTTTAAGAACGACCTATCAATTTGC
TCAATCTCGCGTTCTTTATCAGCGACCTATGTTGAGCATGCCAACAGTATCACAAGGATTAG
CGCGCAGTTTTGCTCGGCTTTTACTGACTGAAATTGCCGCATTGGTATGCAGCAAAGCAATT
GCCTATGCTGCGCCTAGTTTGAGTGTGTATTCGGCATTTTGTAAGTATTGGGTGCCTAGGCA
GACTGAAATAGCGATTGATGAGTTAAAAGTAATCTTAGGTGCTCGCTATTACCTGCGTGATG
AACACCAATTTGGTATTTTCCAAAAAATGCTGCGAGACAATGCCGTGGTATCGTTATTTGAT
GGTAGCAGTCAGGTCAATCTGGCACTCATTGCCACACAAACCAATGCCCTTGCCAGCAAATT
ATTGGCAGCACAAGAGACAGTCCATGATGTTGCCAAGTGGTTTAGTTTAACCAGCGACCAAG
CATCAAGTGCTATCCAAGCGGATGAATTAAAACTTAATAATCTTGGCCAAGACCCAATTCTA
GCAAGTTTTATTGCTGTGTGTGATTCAAACCTGATCACGACTCAAAGTGATCGTTTGCGGGC
GCTTTATGTTGAATTAAAACAGCGAATTCTTCGCTGGTGTGATGAGATAATAACGCTCAAAG
AGCAAGAGCTAAATGATCCCAGCAGTGTGGTGCGATTTGCCAAAGCCAAACAATATACCCAA
TTGTTAGCCGCGACCTGTTGCGCCTTAACCATTATCCATAATCCAAGCCATTTGCTGTTATC
AAATGATGATATTTCTAATGATTTGATGTCATTGCTCCTCAATGACACTGACCAATGGCAAC
CTTGCGAAGCGTTGTTGCAAGCATTCGCACAACAATTTGAGACGCAAAAATGGTTTTCACTT
TTGCCAATATAA

>TamT

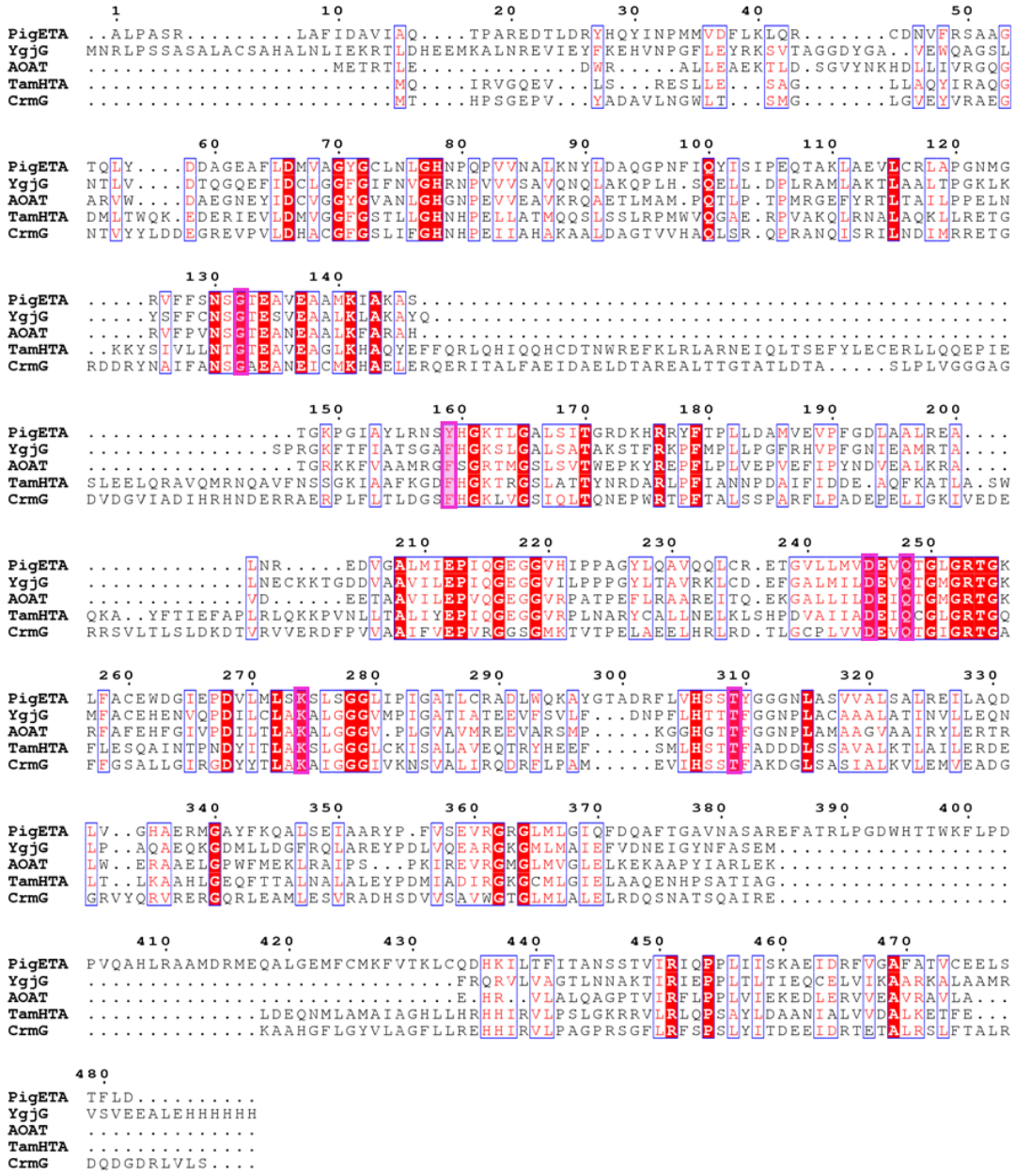
MHDDNTFYAAAAKLEQQCGDPSDEQQFINFKTSSLLDETQQFPTAYIEYLQQLGYFEYFVPS
EQGGQLKSFAQLLLICRALSRRDLNAAIATGQTMLGALPVWLCGSSAQKNQLASHILQGHLG
CLALTEKRHGSNLLASEVSATATATGYELSGEKWLINNATKGHTMTLLARKQETSGRSELAL
LFIDKTNLTEFENLDKINTHGIRCADISGIRFNNTEVASDVVIKTAEPGIYGVMKALQISRI
LCAGFSLGALDTCLRTTYQFAQSRVLYQRPMLSMPTVSQGLARSFARLLLTEIAALVCSKAI
AYAAPSLSVYSAFCKYWVPRQTEIAIDELKVILGARYYLRDEHQFGIFQKMLRDNAVVSLFD
GSSQVNLALIATQTNALASKLLAAQETVHDVAKWFSLTSDQASSAIQADELKLNNLGQDPIL
ASFIAVCDSNLITTQSDRLRALYVELKQRILRWCDEIITLKEQELNDPSSVVRFAKAKQYTQ
LLAATCCALTIIHNPSHLLLSNDDISNDLMSLLLNDTDQWQPCEALLQAFAQQFETQKWFSL
LPI

Appendix 25: The *tamH* gene sequence (GenBank: EAR29362).

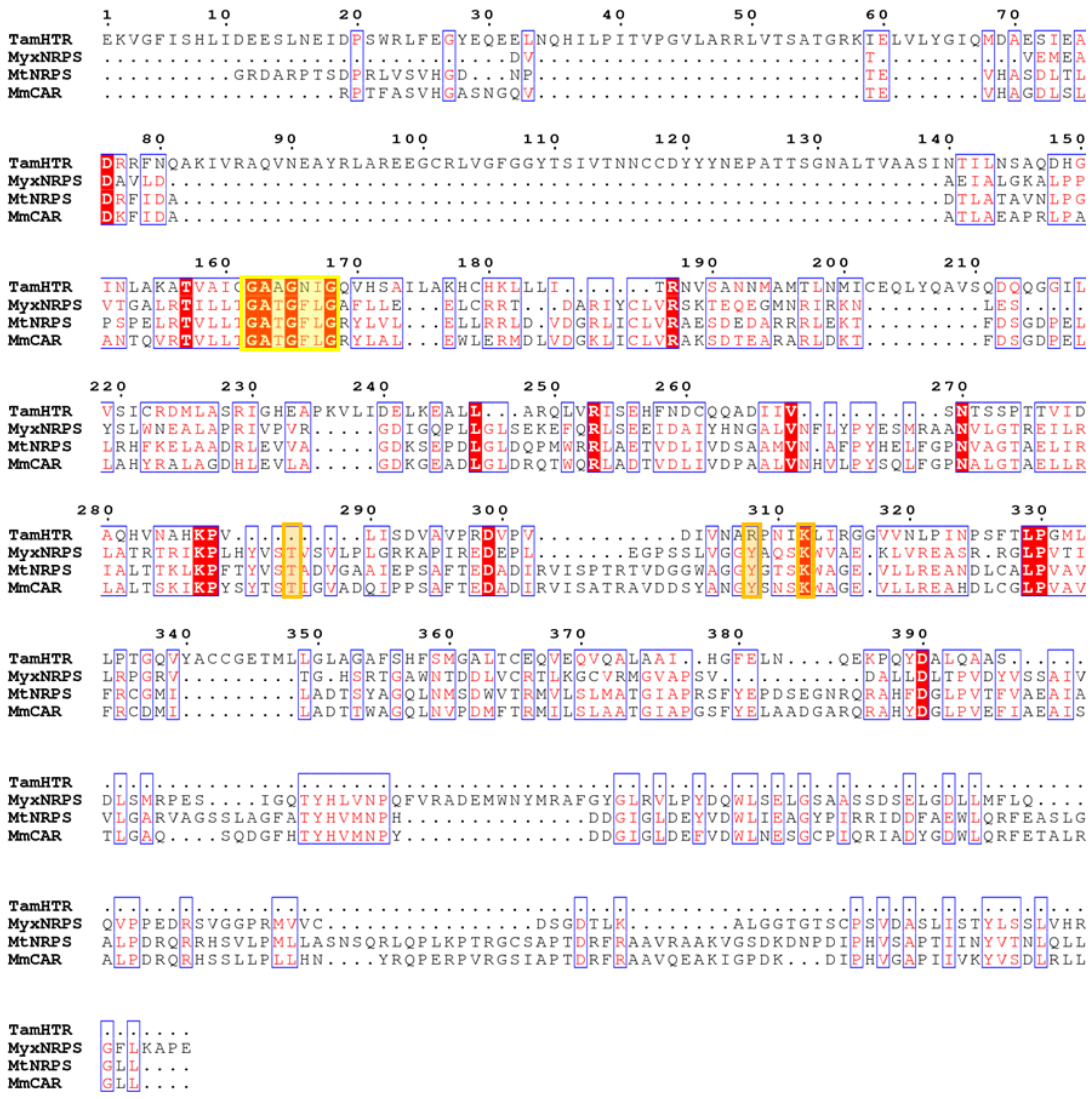
ATGCAAATTCGTGTTGGACAAGAAGTTTTAAGCCGAGAATCACTGTTAGAAAGTGCAGGTTT
ATTGGCTCAATATATCCGTGCGCAAGGCGATATGTTGACTTGGCAAAAAGAGGATGAGCGCA
TTGAAGTACTGGACATGGTCGGTGGCTTTGGCAGTACCTTATTGGGTCATAATCATCCAGAA
TTGCTTGCAACTATGCAGCAAAGCCTCAGTAGTTTAAGGCCAATGTGGGTGCAAGGTGCTGA
GCGACCTGTAGCAAAACAATTACGCAATGCACTTGCGCAAAAGTTGCTACGAGAAACAGGCA
AAAAATACAGTATTGTTTTACTCAATACTGGCACTGAGGCAGTAGAAGCTGGATTAAAGCAT
GCGCAATATGAGTTTTTTCAGCGTTTGCAGCACATTCAACAACACTGTGATACCAACTGGCG
AGAATTTAAATTGCGTCTCGCACGTAATGAAATCCAGCTCACTTCAGAGTTTTATCTTGAGT
GTGAGCGGTTATTACAGCAAGAGCCTATTGAAAGTCTTGAAGAGTTGCAGCGTGCAGTGCAA
ATGCGTAATCAAGCCGTGTTTAATTCATCGGGTAAAATAGCCGCATTTAAAGGCGATTTTCA
TGGTAAAACCCGAGGCAGCCTTGCAACGACTTATAATCGCGATGCACGTCTGCCATTTATTG
CCAATAATCCTGATGCAATATTTATTGATGATGAAGCGCAGTTTAAGGCGACCTTAGCAAGT
TGGCAAAAAGCCTATTTTACCATTGAGTTTGCGCCATTACGTTTGCAGAAAAAGCCTGTAAA
TTTGTTAACTGCGCTAATTTATGAACCCATTCAGGGTGAAGGCGGTGTGCGCCCACTTAATG
CGCGTTACTGTGCTTTATTAAACGAGCTTAAACTCAGTCATCCTGATGTCGCGATTATCGCA
GATGAAATTCAATGTGGTTTAGGTCGTACTGGGCAATTTTTAGAAAGCCAAGCGATAAATAC
ACCCAATGATTATATTACGTTGGCAAAATCACTGGGTGGTGGACTGTGTAAAATCTCCGCTC
TCGCTGTGGAGCAAACTCGTTACCACGAAGAATTTAGCATGCTCCATAGCACCACTTTTGCT
GATGATGATCTCAGCAGTGCGGTCGCGCTAAAAACACTGGCAATTTTAGAACGTGATGAGTT
GACCTTAAAAGCCGCACATTTAGGCGAGCAATTTACGACAGCACTCAATGCGTTAGCACTTG
AATATCCCGATATGATAGCGGATATACGAGGTAAAGGCTGTATGCTTGGCATCGAGCTTGCT
GCCCAAGAGAATCATCCAAGCGCCACAATAGCAGGCTTGGATGAGCAAAATATGCTGGCAAT
GGCGATTGCTGGACATTTGTTACATCGTCATCATATTCGTGTATTGCCTTCATTGGGGAAAC
GTCGAGTATTGCGTTTGCAACCTTCAGCTTATTTGGATGCCGCAAATATTGCGTTGGTCGTT
GACGCACTAAAAGAGACCTTTGAGCTTATTCGTCATCATCATGTTGCAACTTTGCTTGCGCA
TCTTATTCATACCGATAAACCACACAGCGCATTTGCAACGGCTTATCAACATCCAATTCGTG
AAGAGGCAATTCCTGCTGGTGTCGAAAAAGTAGGTTTTATCAGTCACTTAATTGATGAAGAA
AGCTTAAATGAAATAGACCCTAGTTGGCGTTTGTTTGAGGGCTATGAGCAAGAAGAGCTTAA
TCAACATATTTTACCTATTACCGTCCCGGGTGTTTTGGCGCGCAGGTTAGTCACCTCTGCTA
CCGGTCGTAAAATTGAGCTCGTTTTGTATGGTATTCAAATGGATGCTGAGAGCATTGAGGCC
GACCGTCGCTTTAATCAAGCTAAAATTGTGCGAGCTCAAGTCAACGAAGCGTATCGCTTAGC
CCGTGAAGAAGGCTGCCGTTTAGTTGGATTTGGCGGTTATACCTCAATCGTCACTAATAACT
GCTGCGACTATTATTACAACGAGCCTGCTACAACGTCAGGTAATGCGCTGACAGTTGCGGCC
AGTATTAATACTATTTTAAACAGTGCGCAAGATCATGGTATCAACCTCGCAAAAGCAACGGT

CGCTATTTGTGGCGCCGCTGGCAATATCGGCCAAGTACACAGCGCGATATTAGCAAAACATT
GCCATAAGCTGCTGTTAATTACTCGTAATGTCAGCGCGAATAATATGGCAATGACCTTAAAT
ATGATTTGTGAGCAGTTATATCAAGCAGTGAGCCAAGATCAACAAGGGGGAATTTTAGTCTC
TATCTGTCGAGATATGCTTGCTTCACGCATTGGTCATGAGGCACCAAAGGTCTTAATCGACG
AGCTCAAGGAGGCGTTATTAGCCAGACAATTGGTACGTATTAGCGAACATTTTAACGATTGC
CAGCAGGCTGACATCATTGTCAGTAATACCAGTAGTCCAACAACGGTTATTGATGCGCAACA
TGTTAATGCGCATAAACCGGTACTGATCAGTGATGTTGCCGTGCCCCGTGATGTGCCAGTAG
ATATCGTAAACGCACGTCCGAATATTAAACTTATTCGTGGCGGAGTGGTAAATCTTCCTATT
AATCCGAGCTTTACTCTACCTGGAATGTTATTACCAACAGGTCAGGTTTACGCCTGTTGTGG
CGAGACTATGTTACTGGGTTTAGCGGGGGCGTTTAGTCATTTTAGTATGGGCGCGTTAACCT
GTGAGCAGGTTGAGCAGGTGCAAGCCTTAGCTGCAATTCATGGATTTGAATTGAATCAGGAG
AAGCCGCAATATGACGCCTTGCAAGCCGCTAGCTGA

Appendix 26: Alignment of TamH transaminase (TA) protein sequence with *Actinoalloteichus sp.* WH1-2216-6 CrmG, 5DDS; *E. coli* YgjG, 4UOY; *Thermus thermophilus* acetylornithine aminotransferase (AOAT), 1VEF and *Serratia sp. strain FS14* PigE, 4PPM. The conserved amino acids involved in PLP binding are highlighted in pink.



Appendix 27: Alignment of the TamH thioester reductase (TR) protein sequence with *Stigmatella aurantiaca* MyxA NRPS, 4U7W; *Mycobacterium tuberculosis* NRPS, 4DQV and *Mycobacterium marinum* carboxylic acid reductase (CAR), 5MSO. The conserved GXXGXXG motif for nucleotide binding is highlighted in yellow. The catalytic triad (threonine, tyrosine and lysine) are highlighted in orange but these cannot be identified in the TamH TR sequence probably due to a lack of homology.



Appendix 28: TamA ANL recombinant protein sequence from pEHISTEV expression construct. Tobacco Etch Virus (TEV) protease cleavage removes the residues shown in bold including the 6xHis tag.

MSYYHHHHHHDYDIPTTENLYFQGAMECEASSLIDLLAFDARNKPSQEVFRFVSDNGESEAS
YDYQTLSQEISRIAIGLQALIKTSHNQDQALIVLPQGVQFVTAFYGCMAANVIAVPSFPPKS
QLQIERLQFAITDLGNPIVITNRDILPKLQEHIALDSVRWLLIEDLASVIAQPLSDFRTHEH
SIALLQYSSGTTGKPKGVIITNQNIMENSELIRQSFGHKEDHTRMMLWLPPHHDMGLVGGVM
QGVYTGYPTLLMPTDLFLRSQYRWLKAVSDYRATTTGAPNFAYELAVKNIRESRLAELDLSS
LENLFCGAEPINSHSINQFLDKFAPCGLKPEAFLPCYGMAEATLMVSGKPHGQQYKQLCIDE
PLLKHGMVKPLNTPNAHSLWLVSSGVVHSSLQARIVNPETGTEVAQGQVGEIWLQGSSISPG
YWQDAERTAINFGLPLAGYEETFHRTGDLGFYHQDELFITGRLKEVVIIRGANFYPQDLEYE
TTLAFPELNNCRSAAFSVPKEGKEQLIMAIEVPRNVTEFNQYAKILNGRLVERFGIRADIIL
FLPRKTIKITSSGKLQRVAIKKAYEEQQLPVYFQYQLQGEQIA